

『不動産のための計量分析』レジュメ N0.5

クラス担当教員名※1	学部・学科名	学籍番号※2	氏名※2	提出日

※1:履修登録したクラスの担当教員名を書く ※2:学籍番号及び氏名が未記入のもの、また授業終了後に提出されたものは採点しないので、注意すること。

3. データの散らばり具合と2変数の相関

今回の講義も作業自体は Excel のトレーニングですが、それぞれの指標が持つ意味や考え方はデータ分析の基本です。しっかりと勉強しましょう。

3.1 データの散らばり具合

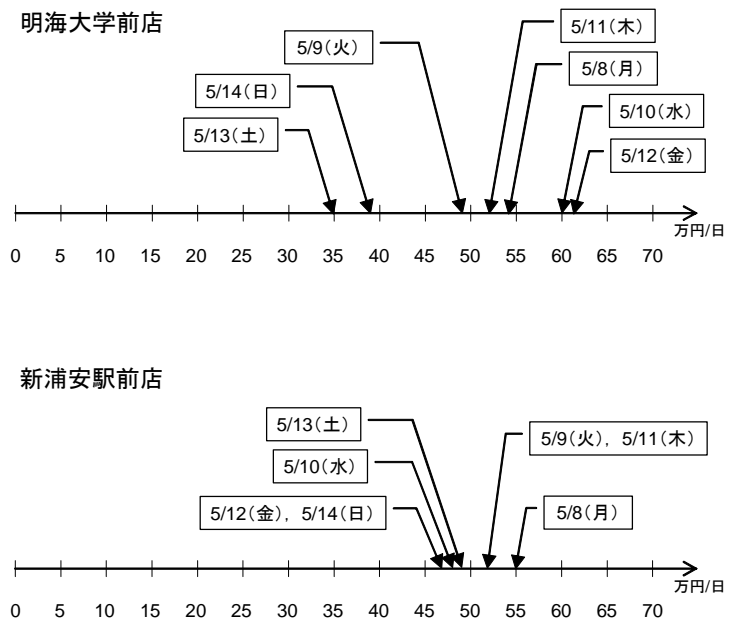
(前回の話の続き) コンビニM&Mの浦安地域マネージャーになったあなたは、浦安にある明海大学前店と新浦安駅前店の一日あたりの売上げを分析することにしました。ある一週間の売上げは下表のとおりです。両店舗の一日あたり売上げの平均値は 50 万円/日と同じです。しかし、両店舗の売上げをよく見ると雰囲気がいぶん違います(右図参照)。明海大学前店のほうが、一日あたりの売上げにムラがあり、売上げデータが散らばっているようです。

表 3-1 両店舗の一日あたり売上げ

	明海大学前店	新浦安駅前店
5/8(月)	54	55
5/9(火)	49	52
5/10(水)	60	48
5/11(木)	52	52
5/12(金)	61	47
5/13(土)	35	49
5/14(日)	39	47
平均	50	50

注)単位は万円/日。

図 3-1 両店舗の一日あたり売上げ



このデータの散らばり具合を表す指標として、**偏差平方和**、**分散**、**標準偏差**があります。**偏差とは、個々のデータから平均値を引いた値**のことで、偏差平方和は、

$$\text{偏差平方和} : \dots(1)$$

で求めることができます。**偏差平方和**は、データ数が多くなると値が大きくなるという欠点があるため、データの散らばり具合を表す指標としてはあまり利用されません。この欠点を解決したものが**分散**で、

$$\text{分散} : \dots(2)$$

で求めることができます。**分散**は、偏差平方和をデータ個数で除して、データ1つあたりに平均化した偏差平方和と言えます。しかし、偏差平方和は、(個々のデータ-平均値)を二乗しているため、もとのデータの単位とは違うものになっています(分散も同じです)。例えば、メートル(m)単位のデータのとき、偏差平方和や分散は平方メートル(m²)単位になり、元のデータと直接比較することができません。そこで、データの散らばり具合を元のデータと同じ単位で比較するために考えられたのが**標準偏差**です。

$$\text{標準偏差} : \dots(3)$$

で求めることができます。**標準偏差**は、分散(データ1つあたりに平均化した偏差平方和)の平方根で、元のデータと同じ単位に戻っています。すなわち、データの散らばり具合を元のデータ単位で表現することができるわけです。

● **偏差平方和、分散、標準偏差は、(i)最小値が0で、(ii)データの散らばり具合が大きくなるほど大きな値となる。**

「不動産のための計量分析」レジュメ N0.5

クラス担当教員名※1	学部・学科名	学籍番号※2	氏名※2	提出日

※1:履修登録したクラスの担当教員名を書く ※2:学籍番号及び氏名が未記入のもの、また授業終了後に提出されたものは採点しないので、注意すること。

作業 1. Ecoome_08_5 をDLして、****data_03 と名前を変えて各自のUF（ユーザーフォルダ）の計量分析の中にセーブしてください。sheet name “コンビニ売上げ” で、両店舗の偏差平方和、分散、標準偏差をステップ①～⑤の順番で求めなさい。****は学籍番号。（2点×6）

3.2 散布図と相関係数

それではコンビニM&Mの新規出店を検討するため、浦安市の需要がどうなるのかを前回講義で話した経済モデルを使って考えてみます。前回検討したように、消費支出（需要）は、

消費関数に関する簡単な仮説：「消費支出は所得に依存するが、**限界消費性向**は1未満の正の数である」

という仮説をたてて、次式のように線形関数に特定化しました。

$$\text{消費関数} : C = a + bY, \quad a \geq 0, 0 < b < 1 \quad \dots(4)$$

表 3-2 家計可処分所得と家計消費支出

年	家計可処分所得	家計消費計	（単位：10億円）					Eその他	(再掲)持ち家の 帰属家賃
			A食品・飲料等	B住居・電気・ ガス・水道	C保健・医療	D娯楽・レ ジャー・文化	Eその他		
1980	156,214	129,673	33,892	25,288	4,895	18,821	46,777	15,903	
1981	167,724	137,919	35,866	27,636	5,709	19,783	48,926	17,221	
1982	176,879	148,088	37,283	29,430	6,630	21,686	53,059	18,461	
1983	184,657	155,821	38,836	31,233	6,881	22,834	56,036	19,735	
1984	194,029	164,024	39,997	33,338	7,060	24,623	59,007	21,159	
1985	203,679	173,563	41,321	35,397	7,680	26,692	62,473	22,716	
1986	210,114	180,542	41,517	36,435	7,513	29,025	66,053	24,204	
1987	215,494	188,879	41,649	38,261	7,564	31,121	70,285	25,931	
1988	228,813	199,021	42,376	40,408	7,407	33,612	75,218	27,607	
1989	244,984	212,758	43,634	43,585	7,259	36,726	81,554	29,994	
1990	264,280	228,287	45,967	47,032	7,272	41,004	87,013	32,401	
1991	282,952	241,350	48,431	50,648	7,334	43,040	91,896	34,979	
1992	292,026	251,248	49,864	54,250	7,418	44,853	94,862	37,578	
1993	296,876	257,462	50,501	57,537	7,461	45,993	95,971	40,174	
1994	303,131	265,639	51,702	60,358	7,904	46,663	99,011	42,173	
1995	304,969	269,280	52,357	62,548	7,885	46,570	99,921	43,785	
1996	305,253	275,291	52,450	64,838	7,955	47,297	102,752	45,289	
1997	311,884	280,601	52,150	67,134	8,680	48,808	103,830	46,798	
1998	314,203	279,447	52,927	68,306	9,240	48,495	100,481	47,955	
1999	311,426	277,365	51,904	69,469	9,537	48,385	98,070	48,941	
2000	306,765	277,160	49,541	70,503	9,937	48,273	98,907	49,910	
2001	297,663	277,440	49,412	71,816	10,159	47,756	98,298	50,998	
2002	298,344	275,356	48,972	72,813	10,441	46,923	96,208	51,878	
2003	297,561	274,246	47,892	73,811	10,900	46,547	95,096	52,719	

計量経済による実証の前に、家計可処分所得と家計消費支出の間になんらかの関係があるのかを**散布図(点グラフ)**で見てください。

作業 2. ****data_03 の sheet name “家計消費支出” で、家計可処分所得と家計消費計の散布図を作って下さい。散布図は縦軸に家計消費計を、横軸に家計可処分所得を取って下さい。（5点）

作業 3. 作業2の散布図を読み取りやすいようにきれいにしてください（グラフが何を表しているかを題目で、縦軸や横軸が何を表しているかが分かるようにしよう）（3点）。

散布図を見るだけでも、家計可処分所得（Y）と家計消費支出（C）の間には関係がありそうなことは読み取れます。**二つのデータの間に直線的な関係が認められるとき、二つのデータの間に相関関係がある**といいます。片方の変数が増加すると他方の変数も増加する傾向にある場合には**正の相関関係**があるといい、片方の変数が増加すると他方の変数も減少する傾向にある場合には**負の相関関係**があるといいます。データの散らばり具合と同じようにどの程度の関係があるのかを指標で表してみます。二つの変数（C、Y）の関係の強さを表す指標として**相関係数**があり、

$$\text{相関係数} : \quad \dots(5)$$

で求めることができます。ここで、分母にある標準偏差は先ほど学んだ方法で計算することができます。分子にあるCとYの**共分散**は、

『不動産のための計量分析』レジュメ N0.5

クラス担当教員名※ ¹	学部・学科名	学籍番号※ ²	氏名※ ²	提出日

※1:履修登録したクラスの担当教員名を書く ※2:学籍番号及び氏名が未記入のもの、また授業終了後に提出されたものは採点しないので、注意すること。

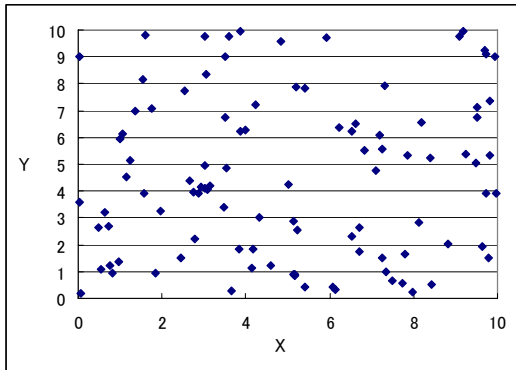
二変数の場合の分散のことであり、二変数を考慮したデータの散らばり具合を表しています。具体的な共分散の計算は、

共分散：

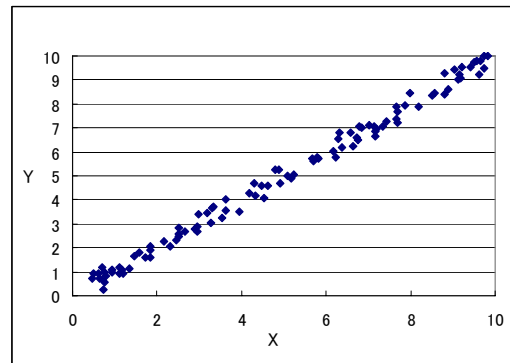
…(6)

で求めることができます。

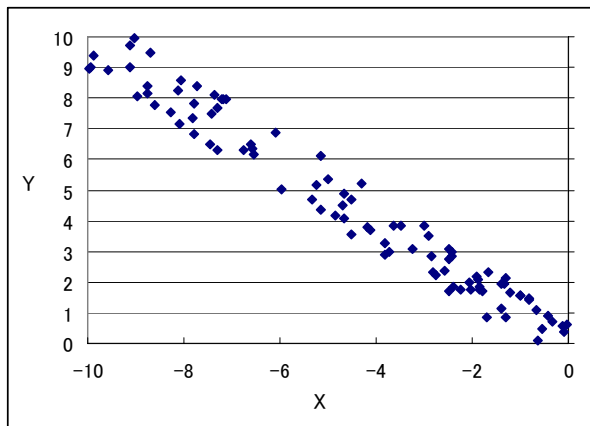
相関係数は $-1 \sim +1$ の間の値をとり、正の値のときには正の相関関係にあり、負の値となるときは負の相関関係にあると読み取ります。また、 $+1$ や -1 に近づくほど相関関係は強くなり、逆に 0 に近づくほど相関関係は弱くなります。散布図と相関係数の関係についていくつか例をみてみましょう。



相関係数=0.083



相関係数=0.996



相関係数=-0.981

『不動産のための計量分析』レジュメ N0.5

クラス担当教員名※1	学部・学科名	学籍番号※2	氏名※2	提出日

※1:履修登録したクラスの担当教員名を書く ※2:学籍番号及び氏名が未記入のもの、また授業終了後に提出されたものは採点しないので、注意すること。

作業 4. ****data_03 の sheet name “家計消費支出（その2）” で、家計可処分所得と家計消費計の相関係数をステップ①～⑨の順番で求めなさい。****は学籍番号。（2点×9）

作業 4 から、家計可処分所得と家計消費計の間にはかなり強い正の相関関係があることが確認できます。つまり、家計可処分所得が増加するほど家計消費支出が増加する傾向にあることが分かりました。

ただし、一般的に相関関係には以下の2点に注意する必要があります。

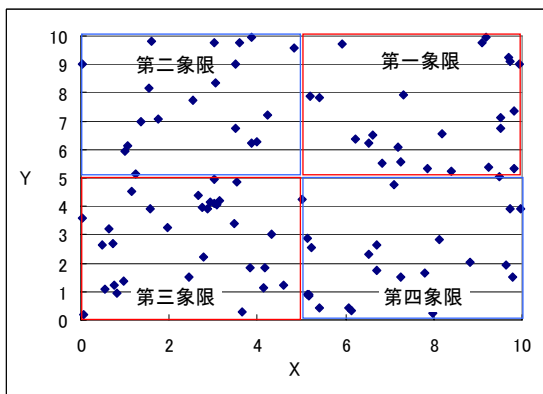
相関係数の注意点①：2変数の相関関係が高くても（相関係数が1に近くても）、その**二つの変数の因果関係は分かりません**。

相関係数の注意点②：経済の時系列データは、それぞれ右上がりの直線トレンドに沿って増加するデータが多く、見せかけの相関が生み出されやすいので注意が必要です。たとえば、日本の高齢者数は、年々増えています。また、日本国内の高速道路の総延長も年々伸びています。この2つの時系列データには相関があります。しかし、高速道路が伸びたから高齢者が増えた、あるいは高齢者が増えたから高速道路が伸びたとは言えないですね。このような相関を**見せかけの相関**と言います。

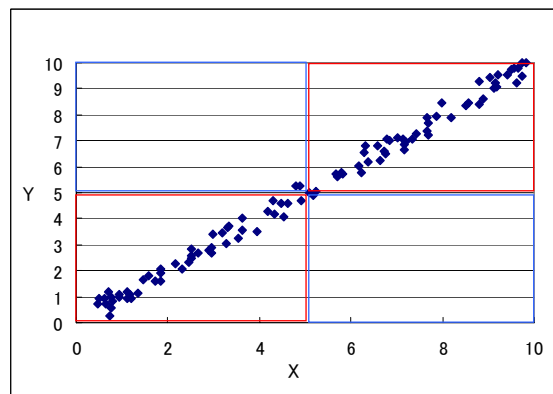
【参考】

共分散を計算するときを使う、偏差の符号は、平均値 (\bar{X}, \bar{Y}) を原点として、第一象限、第三象限のペアではプラスになり、第二象限、第四象限ではマイナスになります。

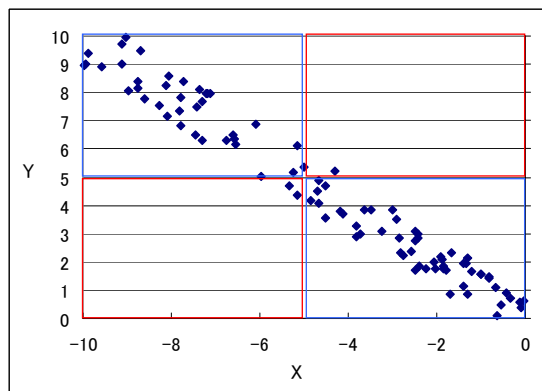
共分散は二つの変数の偏差の積で計算されるため、例えば、相関係数 0.083 のように散らばっていると、プラスの値とマイナスの値が混ざっているため共分散の総和は小さくなりますし、相関係数 0.996 のように右上がりの場合は第一、第三象限が多くなるため総和はプラスの値、相関係数-0.981 のような右下がりの場合は第二、第四象限のペアが多くなるため総和がマイナスの値になることが分かります。



相関係数 = 0.083



相関係数 = 0.996



相関係数 = -0.981