



統計解析って
面白いかも～!

あなたも
計子と一緒に
勉強しようっ!

イッシー、計子と楽しく統計解析!

統計ハニ





イッシー

A大経済学部を卒業の24歳。

現在、広告代理店に勤めている。

広告代理店ということもあり、会社がマーケティングに本腰を入れることになり、経済学部出身のイッシーに期待がかかっている。

本人もビジネスアナリシスに向かって奮闘している。

ビジュアル的にも「イケメン」「さわやか」と評される人気者。

そして、とっても優しい。

だが、そのあたりをいつも妹のみなみに「草食系」と突っつかれる。



計子ちゃん

A大経済学部を志望中の高校2年生。

将来、企業の広報部に入ってマーケティングをやりたいと思っている。

そのため、統計解析を自分でも勉強しているがなかなか進まずにいる。

明るくて元気が良く、とっても前向きでクラスメートの人気者だが、

ちょっと図々しいところ、飽きっぽいところが難点か。。



みなみちゃん

計子のクラスメートであり、イッシーの妹でもある。

学校でもトップクラスの優等生。

一見大人しそうな外見とは正反対にクールで毒舌家。

イッシーも妹とはいえ手を焼いている。

学校では計子ととても仲がよく、テスト前にはいつも計子に勉強を教えている面倒見のいい一面も。意外と姉御肌・・・？！



√ プロローグ



「キンコーンカンコーン・・・」



はあ～あ・・・テストの結果さんざんだったなあ・・・みなみちゃんはいつもの通りバッチリい・・・？



うーん・・・いつもと変わらない感じかなあ。。。



それがいいって証拠じゃないっ！！私なんて・・・このままでA大なんて入れるのかしら・・・



あれ？計子ちゃんってA大志望なの・・・？もう決めてるんだ～。すごいねっ！学部はどこ・・・？



もちろん！経済学部よっ！

企業の広報とかに入ってマーケティングをバリバリやるキャリアウーマン目指すんだもーん！

でも・・・このままだと入れるかすら分からないしい。。

統計解析の勉強もちょっとしてみようと思ったけど何だかチンプンカンプン・・・

誰か教えてくれる人でもいればいいんだけどなあ。。



あっ！そしたら、うちのお兄ちゃんによければ！

うちのお兄ちゃん、実はA大の経済学部出身なの。

大学では統計についていろいろ勉強してたみたいだけど、その後はさっぱりにな

っちゃった

みたいで・・・でも、今行っている会社で必要になって一生懸命勉強し直してるみたいなんだ～。。



え?! そうなの? みなみちゃんのお兄さんってA大経済学部なんだ～?!

やだー!!! 全然知らなかったっ! そりゃ教えてほしいけど、

お兄さんも働いてるし忙しいんじゃない～い???



大丈夫っ! カワイイ子が統計について一緒に勉強したいみたいだよ・・・

なんて聞いたらお兄ちゃん喜んで教えてくれると思うよっ! と言ってもお兄ちゃんがどこまで教えられるか心配だけど・・・



ホントっ?! でも「カワイイ子」っていうのはあんまり強調しないでよ～。

お兄さん期待しちゃうだろうから・・・



お兄ちゃん、計子ちゃんみたいな元気な子好きだと思うよ。

しかもお兄ちゃん、今彼女いないみたいだし・・・



え?! マジでっっ・・・? やだあ!

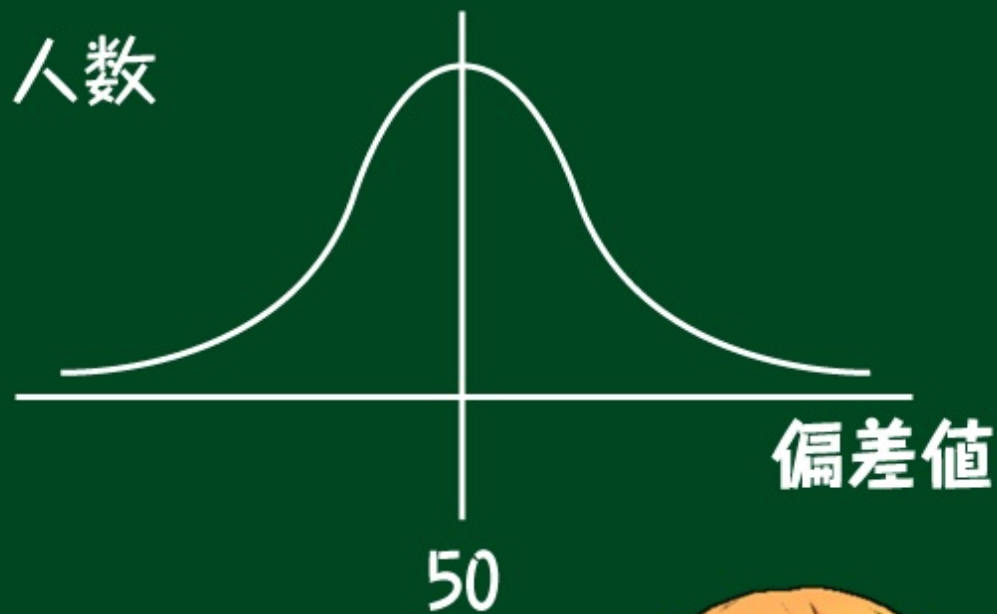
じゃあ、教わっているうちに恋愛に発展しちゃったりしてえ・・・お兄さん格好いい？



う————ん。。。そこは会ってからの楽しみで！
私が言うのもなんだけど好感度はそれなりにあると思うよ・・・



わあっっ！楽しみいっっ！！ みなみちゃんよろしくね～。



chapter1)

√ 偏差値って…?



「ピンポン！」



あ！計子ちゃんいらっしやい！紹介するね。うちのお兄ちゃん！



始めまして！みなみの兄です。いつもみなみがお世話になってます。。。



始めまして！わたし、統野計子といいます！

いつもみなみちゃんに仲良くして頂いていて・・・その他に勉強も教えてもらっちゃってたりい・・・

お世話になっているのはむしろ私の方ですっっ！



話はみなみから聞いてますよ。計子ちゃん、A大の経済目指してるんだってね？



そうなんですっ！

それで・・・統計解析なんぞ勉強しようかと思ったんですが、いまいち・・・よく分からなくてえ・・・



僕も大学には行ったものの、その後さっぱりで・・・

今になって再度勉強しなきゃいけないなっちゃったっていう感じだから、頼りにならない

いかも

知れないけど・・・



いいんですっ！ よろしくお願ひしますっ！！



お兄ちゃんも計子ちゃんと一緒に勉強するって感じよね？



ハハハ・・・そうだね。こちらこそよろしくお願ひします！



計子ちゃん、入って～！



ありがとう！おじゃましま～すっ！



じゃ、あとは2人でどうぞ・・・！ お兄ちゃん、よろしくねっ！



え？！みなみちゃんも一緒じゃないの～？



私はちょっと用があるから！大丈夫！

お兄ちゃんには変なコトしないようにちゃんと言ひ聞かせてあるからねっ！



何だよ、みなみっっ！



フフ・・・お兄ちゃん、草食系だもんね～。じゃあね、計子ちゃん！ゆっくりしてってね～！



ふう～、ホント生意気で困った妹ですよ・・・



ハハッ！みなみちゃん、お兄さんに強いんですね～。お兄さんの話はよく聞いてましたっ！

どんな人かな～って想像してたけど、想像よりカッコいいっ！イケメンですよねっっ？



僕も計子ちゃんの話はよく聞いていましたよ。可愛くて元気があって人気者だって！



え——っ？！ 恥ずかしいっ！ そうだ！ホントお忙しいのにいいんですかぁ？

何か無理やりお願いしちゃったかな～なんて思ってえ。。。



いえいえ、そんなことないです。僕もちょうど今の会社でマーケティングだなんだって急ぎょ統計をまた勉強することになっちゃって。大学以来なもんだから思い出しながら勉強する

感じですよ。

計子ちゃんは高校生なのに統計解析に興味を持ってるなんてすごいよね。

それこそ僕の方がいろいろ教えてもらうことになるかも・・・



やだっ！何を言ってるんですかっ！教わるのは私の方ですよ～。

しかもこんな格好いい先生に教えてもらえるなんて・・・みなみちゃんに感謝ですっっ！



ハハハ・・・そんなことはないですよ。。 それに先生なんて大げさです。



しかも私のあこがれのA大の先輩じゃないですかっ！

計子もこれでA大ま～っしぐら！！っといきたいですっっ！！うふっ。。。



ハハハ……。あ！そうだ！ちょうど、大学の話が出たから早速統計解析について話そうか？

計子ちゃん、「偏差値」ってどういうふうに出すか知ってる？



あ！そうそう、私も「偏差値」ってちょっと疑問があってテストの点数が前より良くなってるのに

偏差値が下がってたりとかあ、あれって不思議じゃないですかあ・・・？



そうなんだよね。いくらテストの点数が良かったとしても全体の中でいい位置につけているとは

限らない。それが偏差値なんだ。

基本的にはテストの点数では計れない、母集団の中での位置を示す指標なんだよ。



母集団・・・？



母集団はテストで言ってみればテストを受けた人達のこと。

計子ちゃん「n」っていう記号見たことある？



あっ！ありまーすっ！！

よくアンケートとかの回答数なんかとイコールで「n」って書かれてますよねっ？



そうそう、あれは母集団の数を表しているんだよ。

アンケートなんかでその母集団がどれだけ有効かどうかというのも統計解析で計れるんだけど、

それはまた後で説明するとして・・・

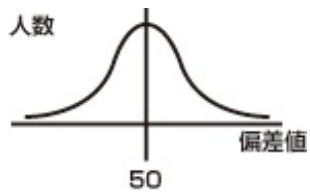
また偏差値の話に戻るけど、偏差値の平均値は常に「50」に設定されているんだ。



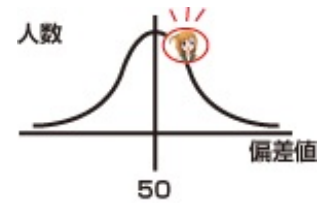
そうなんですかあ・・・？ だからよく偏差値「50」って聞くと普通・・・とか真ん中って言うのねえ。。



そう！その「50」を真ん中にしてそれぞれのバラツキを出すんだ・・・ グラフで見るとこんな感じ！



わっっ！超山型っっ！私はきっとこの辺ね・・・



これは「正規分布」って言って偏差値を出すには理想的な分布図なんだ。
計子ちゃんが今言ったようにこの山型の分布の中にテストを受けた人それぞれの位置があるんだ。



でも・・・理想的・・・ってことはそうでない形もあるってことですかぁ・・・？



そう！そこが偏差値のカラクリ・・・というかテストのカラクリかなあ。。。
こういう分布図になるようにテストを作っているというわけなんだよ・・・



マジでっっ？！ でも・・・ある意味、先生も大変なのね・・・



そうなんだ。テストを作る側がコントロールできるってことなんだよ。。。
というかコントロールしないって言った方が正しいかな・・・
じゃあ、いよいよ本題に入ろうか！実際に偏差値の出し方を見てみよう。



はぁいいっ！



偏差値を出すには「標準偏差」っていうものが必要になるんだ。

ら

これはみんなの点数全体のバラツキを示す指標で、バラツキというのは、要は平均点か

いくら離れているかということなんだよ。

「標準偏差」を求める前にまずは一人一人の平均点との差「偏差」を求めてみるよ！

例えば・・・Aさんの点数が74点でテストの平均点が67.25点だったとすると、

その偏差は $74 - 67.25$ で6.75になるんだ。ここで、全体のばらつきの量を考えてみよう。

AさんからTさんまでの偏差の合計を出せば、それが全体のばらつきの量ということになるよね。

でもこのまま偏差の合計を出すと0になってしまうんだ。。。。

ここに実際のエクセルデータがあるから、計子ちゃん計算してみない？

	得点	偏差
A	74	6.75
B	64	-3.25
C	67	-0.25
D	57	-10.25
E	60	-7.25
F	86	18.75
G	87	19.75
H	86	18.75
I	56	-11.25
J	57	-10.25
K	37	-30.25
L	61	-6.25
M	47	-20.25
N	46	-21.25
O	62	-5.25
P	60	-7.25
Q	100	32.75
R	65	-2.25
S	91	23.75
T	82	14.75



あらっ！ホントだっ！



偏差は平均値との差だから、偏差の合計は必ず0になるんだ。

けど、このままだと全体のばらつきの量がわからないね。そこで偏差を2乗するんだ。

2乗した偏差を足し合わせれば、全体のばらつきの量が出せるよ。

6.75^2 (Aさんの偏差の二乗) + 3.25^2 (Bさんの偏差の二乗) + 0.25^2 (Cさんの偏差の

二乗)

+ 14.75^2 (Tさんの偏差の二乗) = 5293.75 (偏差平方和)



さすがに二乗すると数字が大きくなっちゃいますねっっ・・・！



そうなんだ。でも大丈夫。。。今度はこの数字を比較しやすいように小さくするよ。
テストを受けた人数でこの5293.75という「偏差平方和」を割るんだ。。。

5293.75 (偏差平方和) \div 20 (テストを受けた人数) = 264.6875 (分散)

これが「分散値」というんだ。

「分散」とはデータが平均値を中心にして、どのくらいばらついているのかを示した数値のことを

言うんだよ。



あれ？！確かに「5293.75」から明らかに小さくはなったけど、まだこれだとテストの点数より

大きくなって何だか使えない数字って気がするんだけどお・・・



そう！計子ちゃん、その通り！！さっき「偏差平方和」ていうのを出したよね？
あそこでそれぞれの偏差を二乗してしまっているから元に戻さないといけないんだ。。。

分散値の平方根をとるってことだね。。。 $\sqrt{264.6875}$ (分散値) = 16.269 (標準偏差)

そう、この16.269という数値がさっきから話している「標準偏差」というもので、
今回のテストの点数のばらつきを示す指標になるんだ。



なるほど！これが「標準偏差」なのね！「標準偏差」いただきましたあっっ！！

chapter2)

✓ 偏差値を求めよう!





計子ちゃん、次はいよいよ「偏差値」の求め方に入るよ！



わっ・・・！とうとう本題ね・・・



まずは、さっき (chapter1) も出てきた「標準偏差」と「偏差」を使って「標準化」という作業を行うんだ。。



標準化～・・・？ 何だか分かるような分からないような・・・



おそらく計子ちゃんが考えているイメージで合っていると思うよ。

標準化というのは得点の行列の標準偏差を1にすることなんだ。

標準化後の平均値を「0」、そして標準偏差を「1」にすることで、だいたい自分の点数が

どのくらいのレベルにあるかが分かるんだ。

求め方は、 $\text{偏差} \div \text{標準偏差}$ だよ。実際に標準化した表を見てみよう。

	得点	偏差	標準化
A	74	6.75	0.414893907
B	64	-3.25	-0.199763733
C	67	-0.25	-0.015366441
D	57	-10.25	-0.630024081
E	60	-7.25	-0.445626789
F	86	18.75	1.152483075
G	87	19.75	1.213948839
H	86	18.75	1.152483075
I	56	-11.25	-0.691489845
J	57	-10.25	-0.630024081
K	37	-30.25	-1.85933936
L	61	-6.25	-0.384161025
M	47	-20.25	-1.244681721
N	46	-21.25	-1.306147485
O	62	-5.25	-0.322695261
P	60	-7.25	-0.445626789
Q	100	32.75	2.01300377
R	65	-2.25	-0.138297969
S	91	23.75	1.459811895
T	82	14.75	0.906620019
	平均 67.25	標準偏差 16.2692194	標準偏差 1



あっ！ホントだっっ！！

マイナスが平均より下、プラスが平均より上って感じにちゃんとなっ
てるっ・・・！



そうなんだ。

この表で見ると100点を取ったQさんの標準化得点は2.01....

37点のKさんの標準化得点は-1.859....

ちょっとマイナスだけど67点を取ったCさんが一番平均値に近いって感じだよな！



ホント、ホント～！だんだん偏差値の姿が明らかになってきてますねっ！

でも、標準化って何だかありがたいって感じい～！

だって、偏差と比べるとそれぞれの数字が大きくないからテストの点数良なくても

あんまり目立たないしい・・・2.01と-1.859で比べてもそんなに大差ないって感じだも

～ん。。。



へ・・・?! 計子ちゃん、そういう見方・・・?!



あれ?! ダメですかあ・・・?



いや・・・いいと思うよ。。。やっぱりみなみが言うとおりの面白い・・・



へ・・・?!



(気をとりなおして・・・) そうだ! さっき、計子ちゃん「値が小さい・・・」って言ってたよね?

次は値の見栄えを良くするために値を10倍にして標準偏差を10にするよ。
実際の表がこれだよ。。。

	得点	標準化	×10	+50
A	74	0.414893907	4.148939	54.14894
B	64	-0.199763733	-1.99764	48.00236
C	67	-0.015366441	-0.15366	49.84634
D	57	-0.630024081	-6.30024	43.69976
E	60	-0.445626789	-4.45627	45.54373
F	86	1.152483075	11.52483	61.52483
G	87	1.213948839	12.13949	62.13949
H	86	1.152483075	11.52483	61.52483
I	56	-0.691489845	-6.9149	43.0851
J	57	-0.630024081	-6.30024	43.69976
K	37	-1.85933936	-18.5934	31.40661
L	61	-0.384161025	-3.84161	46.15839
M	47	-1.244681721	-12.4468	37.55318
N	46	-1.306147485	-13.0615	36.93853
O	62	-0.322695261	-3.22695	46.77305
P	60	-0.445626789	-4.45627	45.54373
Q	100	2.01300377	20.13004	70.13004
R	65	-0.138297969	-1.38298	48.61702
S	91	1.459811895	14.59812	64.59812
T	82	0.906620019	9.0662	59.0662
	平均67.25	標準偏差 1	標準偏差 10	標準偏差 10



うーん。。。私的にはさっきの方が・・・



また、そこ・・・?!

でも見やすくなって、だんだん偏差値に近づいてきた感じでしょ・・・?



もしかしてえ・・・これに「50」を足せば「偏差値」になるとかぁ・・・?



計子ちゃん、その通りっ！

最初に話した偏差値の平均値「50」を足せばこれで偏差値は完成！

式はこれだよ！ $\text{偏差} \div \text{標準偏差} \times 10 + 50 = \text{偏差値}$

（しかもって計子ちゃん、実はものすごく分かっててオレがからかわれてるとか・・・）



え?!何か言いましたっ?



いやいやいや・・・分かってくれたらいいな～って。。。



説明が分かりやすいからバッチリですよお～!! さすがセンセイっ!



ハハハ・・・ありがとう。。。 (やっぱり上手か・・・?!)

コラム（偏差値が指標となる時、ならないとき）



計子ちゃんも疑問に思った山型の正規分布。

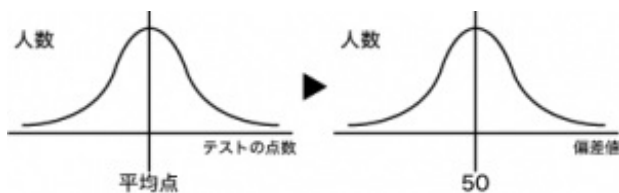
これが山型ではない場合は一体どうになってしまうのでしょうか・・・？

偏差値はある条件を満たしている時のみ利用価値のある値です。

条件を満たしていない時は、指標としてはあまり役に立たないのです。

条件・・・母集団(得点分布)が正規分布に近いこと。

正規分布とは平均値付近が最も高い値とる左右対称の鐘型分布のこと。



母集団が正規分布であるとすれば、偏差値の値によって母集団の中での位置を正しくつかむことができます。

偏差値50以上（あるいは50以下）は、全体の50% 偏差値60以上（あるいは40以下）は、

全体の15.866% 偏差値70以上（あるいは30以下）は、全体の2.275% 偏差値80以上

（あるいは20以下）は、全体の0.13499% 偏差値90以上（あるいは10以下）は、全体の

0.00315% 偏差値100以上（あるいは0以下）は、全体の0.00002% 上記から、例えば

偏差値60の大学試験を突破するためには、全体の上位15%以内に入ればよいということに

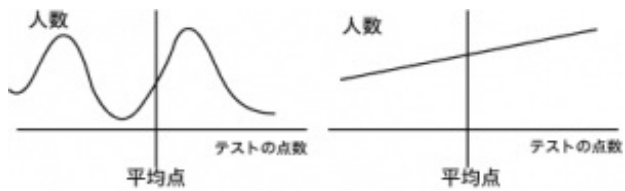
なりますね。自然現象には正規分布に従う事象がいくつか見られます。

たとえば身長分布などがそうです。

しかし、テストの得点分布が必ずしも正規分布に従うわけではありません。

したがって、テストの作り方が大事になります。得点分布が2つのピークを持つような場合や、

平坦な場合は偏差値から上記のような結果を導くことはできません。



上記のような得点分布では、偏差値は指標とはならないのです。

試験によって、得点分布が大きく異なっているのが分かるはずです。

一般に模擬試験では、得点分布が掲載されているので、これを参考にするとよいでしょう。



なるほどね～。正規分布であることが前提での偏差値ってワケねっ！

まとめ（偏差値を操作する・・・?!）



計子ちゃん、正規分布についてとっても納得してくれたみたいだね・・・！



もちろんでーすっ！

コラムに書いてあったから、思わず模擬試験の得点分布も調べて見てみちゃいました～っ！

こういったものが公開されているんですねえ～。



そうなんだ。予備校なんかが出しているセンター試験の得点分布表を大学別に比較して見ても面白いかもね。。。

正規分布になるように先生達もテストを作っているって話がさっき出たと思うけど、ここで計子ちゃんの気になっている「大学偏差値」の話をするね。



「大学偏差値」って何か違うんですかぁ・・・？



今までの説明で偏差値っていうのは同一試験内での指標ってことは分かったよね？



はい・・・



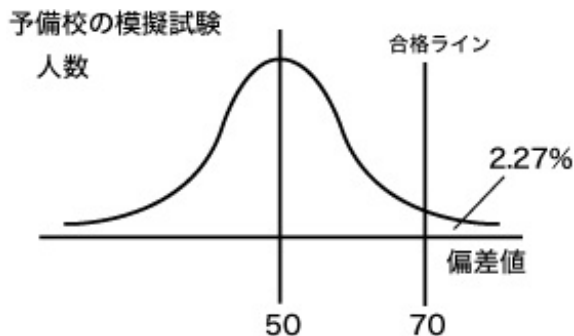
ただそうすると、大学によって個別の試験を実施するという状況では大学の偏差値は決まらないんだよ。ここが高校の偏差値と違うところかな。



じゃあ、どうやって決めているんですかぁ・・・？



基本的に大学の偏差値は予備校の模擬試験によって求められているんだ。
模試の結果と大学合格者の追跡調査によって合格ラインの偏差値が決まるんだ。
一般的には大学合格者の模試偏差値の平均値を求めているようだね。
だから、予備校間で偏差値が異なっているんだ。



例えばこの図のように、ある予備校の模擬試験での合格ラインが偏差値70の場合は、受験生全体の2.27%が合格するということだよ。



から

なるほど～・・・予備校である程度の偏差値をとっていても合格ラインはそれぞれ違う
そこもちゃんと見ておかないといけないってことねえ・・・。
勉強さえやっていればいいってもんじゃないんですねえ～。。。
大学受験するのもいろいろと大変だわ～。



うよ。。。

大変かも知れないけど、それを知ってるか知らないかで受験にも大きく差が出ると思うよ。。。
計子ちゃんは得したと思うけどな～・・・



はあ～・・・



あ！そうそう、もっと面白い話するね！



もっと・・・？！



ズバリ、大学偏差値を上げる方法！！



偏差値を上げる・・・？！ 正規分布にするためにテストを作るっていうのは分かったけど・・・

上げるって・・・？！ 偏差値を・・・？！ですよねぇ～・・・???



そうなんだ。

実は実際に大学で外見上のレベルを上げるために偏差値を操作することがあるんだ。

その操作にはこういった方法があるよ。

1.定員を減らす

推薦入試枠や内部進学枠を拡大し一般入試枠の定員を減らすことによって、合格ラインの

偏差値を上げる。または試験日程を増やすことにより定員を減らす。

現在、偏差値操作ないし他の目的で、AO入試や推薦入試が拡充されており、2009年の春に

私立大学を受験した人のうち一般入試を受験したのは、『読売新聞』の調査では44%であり、

何らの学力審査を経ないで大学生になる人は相当数に上る。

ゆえに、大学生の学力低下の一因との指摘がある。 (wikipedia 偏差値操作より)

2.科目数を減らす

科目数が少ない場合は、特定教科だけ得意とする学生が偏差値を押し上げるという傾向が

ある。



えー・・・！！ 推薦入試にはそんな背景があったのね～。

AO入試は実際の試験がなくって個性や意欲が面接などでアピールできるって確か聞いたことあるけど、私はそっちかしらあ～・・・



も

AO入試はここ最近、入学した学生の学力が下がってきているとかで取りやめる方向に

なっているそうだよ。計子ちゃん、ここはバッチリ試験を受けて合格しようよ！
喜びも違うぞ～！！



か～・・・？

そうなんだ・・・ 何だかお兄さん、予備校のCMの人みたいになってません



だよ。

あ、ゴメンゴメン・・・違うキャラだった・・・？？？

とにかく、偏差値っていうのは今の時代は絶対的なものではなくなってきているん

のが

偏差値の下落を大学側が警戒することで、軽量入試での入学者が大幅に増えてきている

原因だよな。大学の偏差値は意味を失いつつあるってことかな・・・



え?! もしかして、お兄さんってAO入試じゃないですよ～・・・？



何・・・?! 違うよ! 僕はちゃんと試験を受けてA大に入ったんだよー!



のにい・・・

あ～・・・そうなんだ。。AO入試だったらそのこといろいろと聞こうかと思った



え・・・?! 計子ちゃん、AO入試にシフトしちゃってるわけ～・・・?!

せっかく統計解析の勉強始めたんだから、試験受けようよ・・・



あ・・・ハハハ！そうだった！！　じゃ、AO入試は計子の次の策で行きまーすっ！



諦めてないんだ・・・　ふう。。。

chapter3)

✓ あなたは文系？それとも理系？



文系？

理系？



ただいま～！



あ！みなみちゃんお帰りなさ～い！



お帰りっ！



計子ちゃん、どう？ うちのお兄ちゃんに、ちゃんと教えてもらえた～・・・？



もうバッチリだよっ！偏差値のこととかすごい勉強になっちゃったあ・・・！



ホントー？！ 良かった～！ こんな兄でも頼りにしてくれるなんて・・・
お兄ちゃん良かったねっ！！



おいっ！みなみ！！



ハハハ～！



あ！そうそう・・・計子ちゃん！ 文理選択ってもう決めた・・・？



あ・・・！理系か文系かってやつう・・・？ 私はもちろん文系よ～・・・
みなみちゃんはもしや・・・？！



うーん。。。理系にしようか悩んでるのよね～。。。



みなみちゃんだったら理系でも絶対大丈夫だよっ！ だって数学とか得意じゃない・・・
？

私いつも教えてもらってるしい・・・



うーん・・・そうなんだけど・・・

まだ計子ちゃんみたいに「これになりたい！」ってものが今のところないんだよね～。

。。

ただ数学がちょっといいからとかで決めちゃっていいのかな～とも思うし・・・
理系の教科全部が得意ってわけでもないし～・・・



あ！それだったらちょっと統計解析で見てみない？



えー？！統計解析で分かるんですかぁ・・・？



うーん・・・分かるというか・・・さっきみなみが言った数学はいいけど他の理系の教科がそうでもないなんて場合に理系を選んでいいのかっていうのには参考になるかと思うよ・・・



えー！！ 是非聞きた〜い！ ね？みなみちゃん！



うん、うん。。。本当に参考になるか心配だけど・・・



自分で悩んでてそこまで言うかっ・・・？！



まあまあ・・・可愛いみなみちゃんのためだと思って下さ〜い。。。私も聞きたいしい・・・！



計子ちゃんにそこまで言われたら・・・仕方ないか〜・・・これも統計解析だしね・・・



よろしくお願いまーすっ！



例えば・・・数学が得意な人は物理や化学にも強いのかとか国語の点数が高い人は社会の点数も高いのかとかそういうのって「相関関係」があるかどうかなんだ。。。



「相関関係」・・・？



「相関関係」っていうのは2つの変数間での共変関係を指すんだ。
この2つの変数というのが例えば化学の成績と数学の成績になるんだよ。。



それって、化学の成績が上がれば数学の成績も上がるってこと・・・？



うん、まあ、全体としてそういう傾向があるかどうかということだね。
でも、これは因果関係じゃないから、数学の成績が上がればおのずと化学の成績も上がる、
と言っているわけではないんだよ。



「相関関係」と「因果関係」は違う・・・？う～ん、分かったような分からないような!?



ははは・・・、まあ、とりあえず相関関係を確かめてみようよ。
それを確かめるためには、何人かの成績データが必要なんだけど・・・



え・・・?! まさか「私達の成績教える・・・!」ってことじゃないよね・・・?



え————っ?!



ハハハハ・・・ まさか! 人の成績はなかなか見せてもらえないよね。
僕が先生だったらいくらでも見れるのにな。



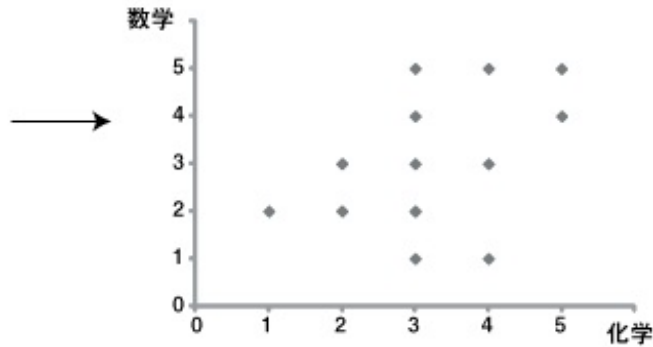
ちょっとお兄ちゃん! 目つきがいやらしいんだけどお・・・!



まあまあ・・・ そんな事だから、今回は僕が作った例で説明してみるよ・・・
早速だけど、この表をちょっと見て。

AさんからNさんの化学と数学のそれぞれの成績なんだけど、これをこんなグラフに置き換えて
みるよ。

	化学	数学
A	4	3
B	5	5
C	3	5
D	4	3
E	3	3
F	1	2
G	2	2
H	2	3
I	3	1
J	3	2
K	5	4
L	4	1
M	4	5
N	3	4
平均	3.285714	3.071429



X軸に化学の成績、Y軸に数学の成績を入れてみたよ。この図のことは「散布図」って言うんだ・・・

この図を見てどう思う・・・？



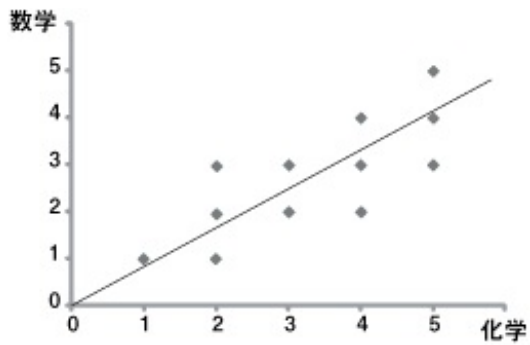
うーん。。。一見、右肩上がりのものもあればそうでないものもあるっていうかあ、何かばらついてるって感じかなあ・・・



そうねえ・・・何かこの図だけでは判断できない・・・って感じがする。。。



そうだね。じゃあ、この図だったらどう・・・？



あっ！これだったらすごく右肩上がりって分かるっ！



うん、うん・・・この分布だと化学の成績がいいと数学の成績もいいんだなって思えるわ！



そうだよ。この右肩上がりになっている分布はまさに化学と数学の成績に相関関係があると分かるよね。この右肩上がりになっていることを「正の相関」と言うんだ。

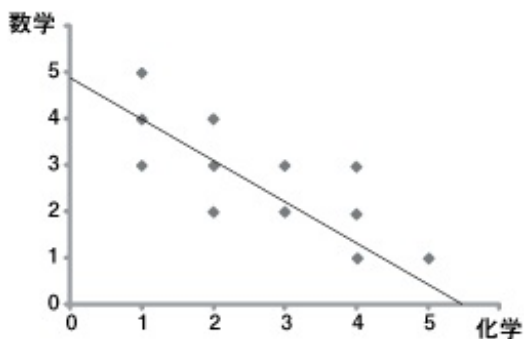


じゃあ、「負の相関」もあるってこと・・・？



そう！こっちの図を見て！これは右に下がってるでしょ？これは「負の相関」と言うんだ。

この図で見ると化学の成績が良くなれば数学の成績は悪くなることになるし、逆に数学の成績が良くなれば、化学の成績が悪くなるってことになるんだ・・・





でも、何かそれってあんまりえない感じい・・・私なんかどっちも良くないしい。。



一般的に考えるとそうだよな。計子ちゃんみたいな人から、どっちも中間、どっちもいって人を対象に統計を取ると「正の相関図」になるのかも知れないし・・・



はあ・・・（何かそれって・・・残念な感じなんですけどお・・・）



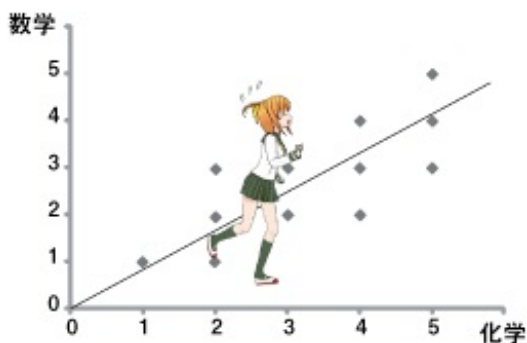
でも、私は化学ってあんまり得意じゃないよ～。
私みたいな人が対象になるともしかして「負の相関図」になるのかなあ。。



そうだね。



でもお、みなみちゃん数学がいいんだからいいじゃな～い・・・？
私なんかどっちも頑張らないとここから脱出できないのよお・・・



でも、計子ちゃん文系って決めてるんだから、文系教科の散布図でいい位置にいれば充分じゃない・・・？



あっ！そっかあ・・・！！それもそうねえ～！



うん・・・！ そうだよ、そうだよ！！ もうっ！お兄ちゃんが変なコト言うから
さ～・・・！



え？！オレのせいなのーっ・・・？！

chapter3-1)

✓ 関係性を数字で表す?!





ハハハ・・・みなみちゃん！毒舌ツッコミはまあその辺にしといて・・・
あたしゃ、大丈夫だから～。。。



そうだよ。計子ちゃん！

せっかく統計解析の勉強してるのに、みなみが余計なこと言うからな～。。。



余計なこと言ったのはお兄ちゃんじゃない？

計子ちゃんの前で兄妹ゲンカしても仕方ないからさっさと話進めて下さーい。。。



分かりましたよー！はあーあ・・・



今日の2人はある意味、「正の相関関係」なのかな～。。。 ???



え・・・?! 計子ちゃん、面白いこと言うね～！

計子ちゃんから相関関係って言葉が出たところで、この「相関関係」について少し突っ込んでみるよ。要は「関係性を数字で表す」ってことかな。。。

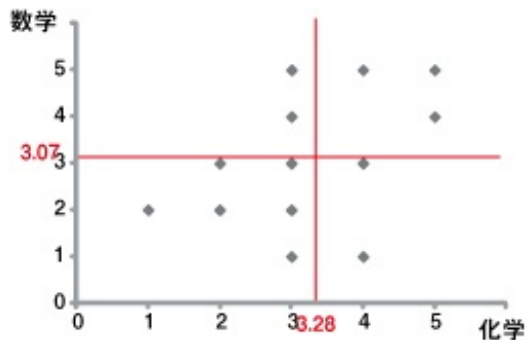


「数字・・・？」ってことはさっきのグラフをまた使うんですかあ～？



そう！今度はさっきの散布図にラインを引いてみるよ。

X軸とY軸それぞれの平均値に引いてみるんだ。



それぞれの点がこのラインからどのくらい離れているかを調べるよ。
 それぞれの点のX、Y座標とその平均値tとの差を求めればいいんだ。
 例えばAさんを例にとると・・・
 $x_A - \bar{X}$ (Aさんの得点-平均点) となるね。
 ということは、それぞれの偏差を求めればいってわけなんだ。

	化学	化学の偏差	数学	数学の偏差
A	4	0.71428571	3	-0.0714286
B	5	1.71428571	5	1.92857143
C	3	-0.2857143	5	1.92857143
D	4	0.71428571	3	-0.0714286
E	3	-0.2857143	3	-0.0714286
F	1	-2.2857143	2	-1.0714286
G	2	-1.2857143	2	-1.0714286
H	2	-1.2857143	3	-0.0714286
I	3	-0.2857143	1	-2.0714286
J	3	-0.2857143	2	-1.0714286
K	5	1.71428571	4	0.92857143
L	4	0.71428571	1	-2.0714286
M	4	0.71428571	5	1.92857143
N	3	-0.2857143	4	0.92857143
平均	3.285714		3.071429	



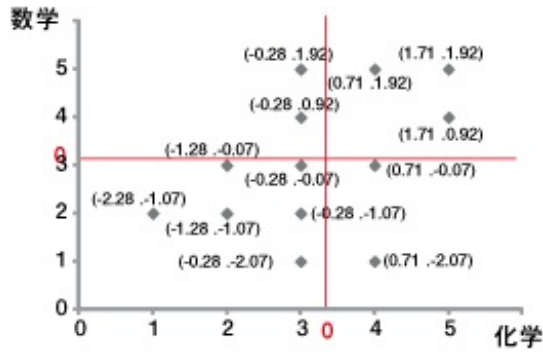
なるほどお・・・



右図でAさんからNさんまでのそれぞれの偏差を出してみたよ。



そしたら今度は全員の偏差を散布図のそれぞれの点に追加してみるよ。



こう見ると、平均点のラインをX、Y軸とすれば、偏差の値はそれぞれの座標点を表しているのが

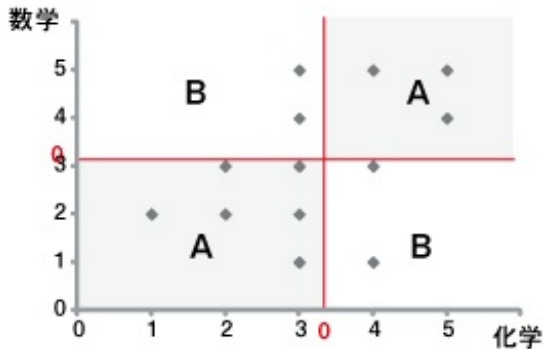
分かるよね？



うん、うん。。。



次はこの散布図を4つのエリアに分けてみるよ。



どうかな・・・？



うーん・・・ これって右上のAのエリアはプラスばかりで左下はマイナスばかり・・・

でもお・・・Bの左上のエリアはプラスもマイナスもあるしい・・・、右下のエリアも何だかおんなじ感じだけどお。。。



あ！ 積がプラスかマイナスかってこと・・・？



ん・・・？！ せき？！



(小さな声で) 計子ちゃん、掛け算よ掛け算・・・！



あー！！ その「せき」！！ね・・・ (いまいち、漢字が思いつかないけどお・・・ま、いっかあ。。。)



そう！ 座標点を見ると一目瞭然だよな・・・！

つまり、Aのエリアは座標点の積が+、Bのエリアは積が-になっているんだ。

ということは・・・Aのエリアが「正の相関」を表すエリアで、Bのエリアが「負の相関」を表す

エリアってことなんだよ。。。



なるほどー。。。



それが「数字で表す」ってことなのね～！



そう！ つまり、AのエリアとBのエリアのどちらが優勢であるかが分かれば、正と負どちらの相関があるかが分かるんだよ！



ほお～！ 分かりやすいわね。お兄ちゃん！



ハハッ・・・ありがと・・・



あらあらあ・・・またまた正の相関関係ってやつう。。。
まさに今、Aの右上のエリアね。。。ふふつ。。。



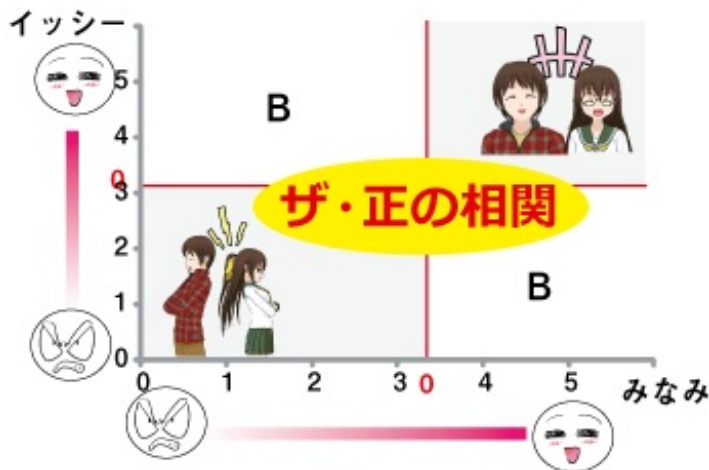
え?! 計子ちゃん、さっきから何の相関関係を例えてるの・・・?



もちろんっ! みなみちゃんとお兄さんの気持ちで一すっ!
さっきはお互い怒っててえ、今はお互い仲良い・・・みたいな。。



それってお互いにいい感情を持っている時は点数が高くて、悪い感情を持っている時は
点数が低い・・・ってこと?!



そうで一すっ!!



あー・・・。 ということね。。。 (小さい声で) やっぱり、計子ちゃんは文系
だな。。

というより文系、理系で当てはめられないってやつか・・・?!



ふふっ・・・計子ちゃんって、ホント面白い！！

chapter3-2)

✓ 相関性を測る・・・?!





ちょっと例えが変わってるけど・・・まあ・・・

計子ちゃん、相関関係についてはとっても理解してくれたみたいだね。。



ほんとですかーっ?! 計子、これからはこっちの路線でいろいろ例えてみようかなっ

!

その方が何か面白いも～ん!



アハッ! いいかも～! 計子ちゃん!



どうぞ・・・ご自由に。。。ふうっ・・・



そうだ! お兄ちゃん、結局、どちらが優勢かはどうやって分かるの・・・?



えーっと・・・(何だか完璧に2人のペースになってるぞ・・・やばいやばい。。。

)

そうそう! さっきの散布図を見て!

それぞれの座標点の積、つまり偏差の積を全て足して合わせてみるんだ。

式はこうだね。

$$(x_A - \bar{X})(y_A - \bar{Y}) + (x_B - \bar{X})(y_B - \bar{Y}) + \dots + (x_N - \bar{X})(y_N - \bar{Y})$$

	化学の偏差	数学の偏差	偏差の積
A	0.71428571	-0.0714286	-0.05102
B	1.71428571	1.92857143	3.306122
C	-0.2857143	1.92857143	-0.55102
D	0.71428571	-0.0714286	-0.05102
E	-0.2857143	-0.0714286	0.020408
F	-2.2857143	-1.0714286	2.44898
G	-1.2857143	-1.0714286	1.377551
H	-1.2857143	-0.0714286	0.091837
I	-0.2857143	-2.0714286	0.591837
J	-0.2857143	-1.0714286	0.306122
K	1.71428571	0.92857143	1.591837
L	0.71428571	-2.0714286	-1.47959
M	0.71428571	1.92857143	1.377551
N	-0.2857143	0.92857143	-0.26531
偏差積和			8.714286

ここで出た値を「偏差積和」というんだ。ここでの偏差積和の値は8.714286。

つまりプラスになったから化学と数学の相関関係は正の相関ということが分かるんだ。



あっ！ まさに、みなみちゃんとお兄さんと一緒だ〜っ！



また・・・そこ・・・?! (いちいち、突っ込まない、突っ込まない。。。)

えーっと・・・この8.714286という値はどのくらいの相関性を表していると思う？



うーん。。。難しいわ〜。だって今いち基準が分からないもの・・・



そうだよね。

例えば相関性を表す値が-1から1までの値をとるとしたら、相関性の度合いを表す尺度として

使えそうな感じはしないかな？



でもお〜、8.714286じゃ、-1から1に入らないですよねぇ。。。



そう！それだけ偏差積和っていうのは、偏差の値によっていくらでも値をとれるってことだから、

現状では尺度として使うことは出来ないんだよ。。。



じゃあ、どうするの・・・？



まず、偏差積和をデータの数で割ってみるんだ。

$$\frac{(x_A - \bar{X})(y_A - \bar{Y}) + (x_B - \bar{X})(y_B - \bar{Y}) + \dots + (x_N - \bar{X})(y_N - \bar{Y})}{n} = 0.622449$$

この0.622449という値が「共分散」というんだけど、これは偏差積の平均値と考えることが

できるんだ。



わっ！ これだったら-1から1の間に入ってますもんね〜っ！



うん・・・ただ、この「共分散」でも相関性の強さを表す尺度としては不十分なんだ。

。。



えーっ?! そうなんですかぁ・・・?



何でかというところ・・・偏差の値の桁数によって値が異なってしまうんだ。。



え・・・?!



例えば・・・今回の化学と数学の成績の単位を100点満点のテストの単位とするよ。
そうすると、こんなふうに偏差積和も3485.714286と桁数が増えてしまうんだ。

	化学(点)	化学の偏差	数学(点)	数学の偏差	偏差の積
A	80	14.28571429	60	-1.428571429	-20.40816327
B	100	34.28571429	100	38.57142857	1322.44898
C	60	-5.714285714	100	38.57142857	-220.4081633
D	80	14.28571429	60	-1.428571429	-20.40816327
E	60	-5.714285714	60	-1.428571429	8.163265306
F	20	-45.71428571	40	-21.42857143	979.5918367
G	40	-25.71428571	40	-21.42857143	551.0204082
H	40	-25.71428571	60	-1.428571429	36.73469388
I	60	-5.714285714	20	-41.42857143	236.7346939
J	60	-5.714285714	40	-21.42857143	122.4489796
K	100	34.28571429	80	18.57142857	636.7346939
L	80	14.28571429	20	-41.42857143	-591.8367347
M	80	14.28571429	100	38.57142857	551.0204082
N	60	-5.714285714	80	18.57142857	-106.122449
平均/標準偏差	65.71428571		61.42857143		3485.714286
共分散					248.9795918



そうなること、共分散の値も桁が増えてしまうってことなのね〜。



そう！



そっかあ！ これだと共分散の値は、14で割って、えっと・・・
248.97959になっちゃうから-1から1の間に入らなくなっちゃうもんね～・・・



そうなんだ。2人とも理解が早くなってきたね～。いやぁ・・・素晴らしい！



それで、どうするの・・・？



あ、えーっと・・・（スルーかよ・・・！）
そうそう！ 計子ちゃん、「標準化」って覚えてる・・・？



え～っと・・・確か、偏差値の出し方（chapter2）の時に出てきたやつですよねぇ・・・



そう！ 偏差値の求め方の時にも言ったけど、標準化というのは得点の行列の平均を0
、
標準偏差を1にすることなんだ。



その標準化を使うことで、-1から1の間の値を出せるってことなのね・・・？



さすが～！ みなみちゃん！！ 聞いてなくてもすぐ分かっちゃう！



そうなんだ。式は前にも言ったように、偏差÷標準偏差。
例えばAさんの化学(x)と数学(y)の成績を標準化すると・・・式で表すとこうだね。
 S_x 、 S_y はそれぞれの標準偏差を表しているよ。

$$\frac{x_A - \bar{X}}{S_x} \quad \frac{y_A - \bar{Y}}{S_y}$$

標準化を実行すれば、どんな値でも平均は0、標準偏差は1になるから、これでデータの桁数を揃えることができるんだよ。



標準化ってすごーいっ！ 標準化、サイコーッ！！



アハハハッ！

$$-1 \leq \text{相関係数} \leq 1$$



chapter3-3)

✓ 相関係数を求める!



標準化最高！と盛り上がったところで実際に計算をしてみるよ。
まずはAさんの化学の点数の標準化をしてみるよ。

	化学	化学の偏差	標準化	数学	数学の偏差	標準化
A	4	0.71428571	0.65094455	3	-0.0714286	-0.0535288
B	5	1.71428571	1.56226693	5	1.92857143	1.44527686
C	3	-0.2857143	-0.2603778	5	1.92857143	1.44527686
D	4	0.71428571	0.65094455	3	-0.0714286	-0.0535288
E	3	-0.2857143	-0.2603778	3	-0.0714286	-0.0535288
F	1	-2.2857143	-2.0830226	2	-1.0714286	-0.8029316
G	2	-1.2857143	-1.1717002	2	-1.0714286	-0.8029316
H	2	-1.2857143	-1.1717002	3	-0.0714286	-0.0535288
I	3	-0.2857143	-0.2603778	1	-2.0714286	-1.5523344
J	3	-0.2857143	-0.2603778	2	-1.0714286	-0.8029316
K	5	1.71428571	1.56226693	4	0.92857143	0.69587405
L	4	0.71428571	0.65094455	1	-2.0714286	-1.5523344
M	4	0.71428571	0.65094455	5	1.92857143	1.44527686
N	3	-0.2857143	-0.2603778	4	0.92857143	0.69587405
平均/標準偏差	3.285714	1.09730654		1.3071429	1.33439584	1

Aさんの化学の偏差は0.71428571で化学の標準偏差は1.09730654だから
偏差÷標準偏差で0.65094455になるね。

これと同じ要領でNさんまでの化学の点数の標準化と数学の点数の標準化を行うんだ。
そこで出た値がさっきの表の標準化の欄に入っているんだ。



うん、うん・・・

ここで、さっきの共分散を求めた式を思い出して！
あの式は偏差積和をデータ数で割ったものだったね。
ここで偏差の値を、いま求めた標準化の値で置き換えてみるんだ。
実際の式はこんな感じ。

$$\frac{\left(\frac{x_A - \bar{X}}{S_x}\right)\left(\frac{y_A - \bar{Y}}{S_y}\right) + \left(\frac{x_B - \bar{X}}{S_x}\right)\left(\frac{y_B - \bar{Y}}{S_y}\right) + \dots + \left(\frac{x_N - \bar{X}}{S_x}\right)\left(\frac{y_N - \bar{Y}}{S_y}\right)}{n}$$

$$= \frac{1}{n} \left\{ \frac{(x_A - \bar{X})(y_A - \bar{Y}) + (x_B - \bar{X})(y_B - \bar{Y}) + \dots + (x_N - \bar{X})(y_N - \bar{Y})}{S_x S_y} \right\}$$

$$= 0.4251$$



ここで出た0.4251という値はズバリ・・・？



-1から1の間っ！



んー・・・！ 確かにそうなんだけど・・・



え・・・？！



「相関係数」と言うんだな～。。。



何言ってるの？！ この流れでいったらどう考えても「-1から1の間」ってなるじゃーん。。。

マジで空気よめてないな～。お兄ちゃん・・・ね？ 計子ちゃん！



確かに～！ お笑いだと相方が聞いてきたことの答えをわざとはずしてウケるってことは

あるけどお、これは完璧はずしてないですもんねえ～。。。



ほんとほんと～！ キャハハッ！



（オ、オレ、お笑いじゃないっ！

しかも、はずしてるのはさっきから計子ちゃんじゃんかよっ！ くう～っっ・・・）



どうしたの・・・？ お兄ちゃん！ そろそろ、この辺でシメのコメントが入るんじゃない～い・・・？



（くう～っっ・・・みなみまでっ・・・！ あくまでもオレは先生、先生・・・）

あははっ・・・そうだ・・・ね。。。

「相関係数」がなぜ相関性の度合いを計れるかというとなんだけど・・・
例えば、化学の成績(x)と化学の成績(x)の相関係数を見ても分かるよ。

$$\frac{\left(\frac{x_A - \bar{X}}{S_x}\right)\left(\frac{x_A - \bar{X}}{S_x}\right) + \left(\frac{x_B - \bar{X}}{S_x}\right)\left(\frac{x_B - \bar{X}}{S_x}\right) + \dots + \left(\frac{x_N - \bar{X}}{S_x}\right)\left(\frac{x_N - \bar{X}}{S_x}\right)}{n}$$



でもお、これってえ・・・相関関係があるのは当たり前ですよねぇ・・・同じものだも～ん。。。



まあ、見てて。。。この式を書き換えると、

$$= \frac{\left(\frac{x_A - \bar{X}}{S_x}\right)^2 + \left(\frac{x_B - \bar{X}}{S_x}\right)^2 + \dots + \left(\frac{x_N - \bar{X}}{S_x}\right)^2}{n}$$

となつて、これは標準化変量の分散を表すんだ。

分散は偏差の2乗和(偏差平方和)をデータ数で割ったものだったよね。

ここでさっき言ったことを思い出して!標準化すると標準偏差は1になるんだったよね。分散は標準偏差の2乗だから、分散も1でしょ。

つまり、化学の成績(x)と化学の成績(x)の相関係数は、標準化変量の分散だから必ず1になるんだ。



だから、最も強い相関は1になるってこと?



そう、化学と化学の成績の相関関係より強い相関性は考えられないから、相関係数は最大で1になるってことなんだよ。

つまり、相関係数(r)は $-1 \leq r \leq 1$ という範囲で推移するってことなんだ。



なるほどね～。。。



深いのね～。統計解析って・・・



(よしっ! オレ、持ち直したぞっ!!)

タイプ診断をしよう！



あなたは文系？理系？それとも・・・芸術系？
あなたのタイプを診断しちゃおう！

タイプ診断で「有名人に例えると・・・」から、何と！「おススメの職業」まで
分かっちゃう！

さーて・・・あなたは何タイプかな～っ・・・！
計子のとっておき画像？！も見られるよ～っ！σ(^-°)

♡ タイプを診断する！ ♡



私たちもタイプ診断してみたよ～。結構、それぞれ違うので是非やってみてね～！

意外な自分が発見できちゃうかもっ。。。

chapter3-4)

✓ 相関係数を散布図で見よう!

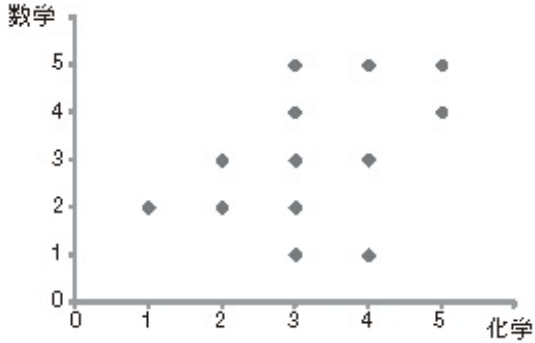




科学と数学の成績の相関係数が0.4251ってことは分かったよね？

さっきも言ったけど、この相関係数は1-から1の間を推移するんだけど、この範囲でどれだけ違うのかを散布図で分かりやすく説明するね。

まずはさっきの科学と数学の成績の散布図はこんな感じだったよね。。

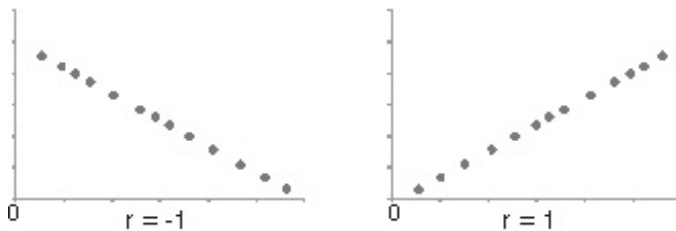


うんうん・・・



じゃあ、他の相関係数を散布図で表してみるよ。

まずは相関係数が-1と1の場合・・・



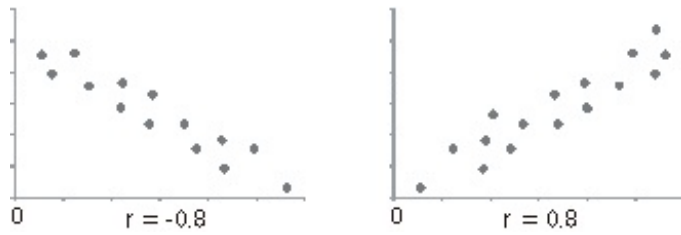
例の右肩上がりと右肩下がりね。



そう！相関関係マックス！ってやつうっ！



うん。これはどうかな。相関係数が-0.8と0.8だね。

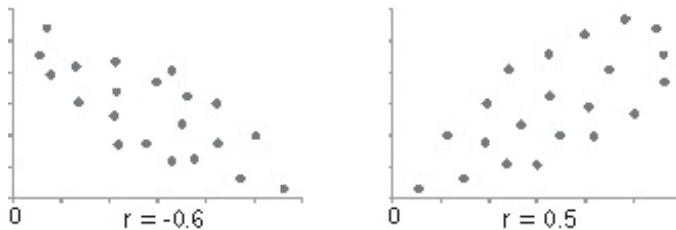


うんうん・・・何となくまだ直線に近い伸び方をしてるから相関関係は高そうね。。。



そう！この位だとかなりの相関性が確認できるんだ。。。

そしたら、これは？相関係数が-0.6と0.5。



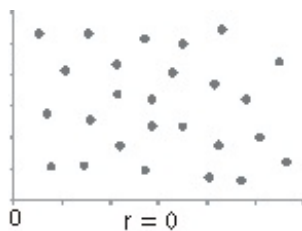
うん・・・何か横に広がってきた～・・・

0.5の方は形で見るとさっきの化学と数学の散布図に一番近いかも・・・



そうだね。この位であれば、まずまずの相関性が確認できるかな。。。

次は最後だよ。相関係数が0。これはどう？



ありゃ～・・・まさに点々バラバラ・・・何か收拾つかないって感じい。。。



そう。まさにその通り！これは相関性がないってことになるね。



でも～っ！



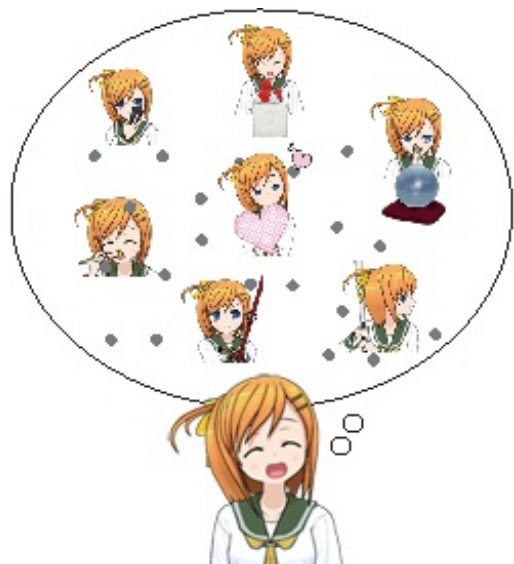
え？！



計子の頭の中ってこんな感じかも～っ！



???



だって、こっち考えてえ、あっち考えてえ、それでよく何だったっけ～・・・？
ってなっちゃうも～ん。。



...



もっと、いいこと言うかと思ったのに。。。まあ、期待したオレが馬鹿だったか～。。。

。

(しかも計子ちゃん、野球と釣りもやるなんて・・・?!)



へ?!何か言いました～?



あ!いやいや・・・



次回はいよいよ統計ハニー初のイベント「タイプ診断」をこの相関係数を使って解析しちゃうよ～! 乞うご期待っ! coming soon!!



ま、まさかの次回予告っ・・・?!

統計ハニー

<http://p.booklog.jp/book/11032>

著者：イッシー

著者プロフィール：<http://p.booklog.jp/users/isssssy/profile>

感想はこちらのコメントへ

<http://p.booklog.jp/book/11032>

ブックログ本棚へ入れる

<http://booklog.jp/item/3/11032>

統計ハニーブログ

<http://i-click.co.jp/toukei/>

電子書籍プラットフォーム：ブックログのパパー (<http://p.booklog.jp/>)

運営会社：株式会社paperboy&co.