

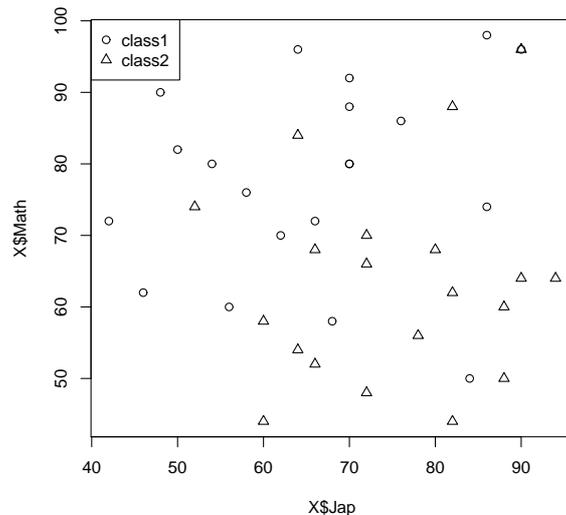
## シンプソンのパラドックス

例 1. 国語と数学の試験の結果を 2 つのクラスからそれぞれ 20 名を無作為に抽出して調べたら、以下のようになった。

(70,92), (84,50), (76,86), (86,74), (86,98), (70,88), (66,72), (68,58), (46,62), (56,60),  
 (70,80), (64,96), (50,82), (42,72), (62,70), (54,80), (90,96), (70,80), (48,90), (58,76)  
 (64,54), (72,48), (82,44), (60,58), (88,60), (72,66), (88,50), (80,68), (82,88), (90,96),  
 (60,44), (94,64), (52,74), (66,68), (64,84), (82,62), (78,56), (66,52), (90,46), (72,70)

このデータの相関係数を調べると、 $-0.0243$ であった。よって、国語と数学の成績に弱い負の相関があることが見受けられる、即ち、国語の成績と数学の成績には関係ないか、むしろ一方の成績がよいものは他方の成績が良くないという傾向が弱いがあること意味している。

しかし、これに疑問をもったある教員が、1組と2組を分けて相関図を作成したところ、それぞれに偏りがあることに気づいた。そこで、それぞれについて相関係数を調べると1組では  $0.2077$ 、2組では  $0.1252$  となり、ともに正の相関がある、即ち、国語の成績と数学の成績には一方がよければ他方の成績も良いという傾向があることがわかった。(この例のデータは筆者がこの現象を示すよう捏造しました。)



このように全体の傾向が、新しい要因を組み込んだとき全面的に否定されてしまうような結果を得ることをシンプソンのパラドックスという。

次は『渡部 洋著 ベイズ統計学入門 福村出版』からの引用です。

例 2. ある病気に対して提案された治療法を調べたところ、左の表のような結果を得た。

ここで、事象  $R$  はその病気から回復するという事象、 $R^c$  は回復しないという事象、 $T$  はその治療法を与えるという事象、 $T^c$  は与えないという事象を表すものとする。このとき、 $T$  の場合の治癒率は  $20/40 = 0.5$  であり、 $T^c$  の場合は  $16/40 = 0.4$  であるから、少なくともこのデータに基づく限りこの治療法を与えるべきとなろう。

	$R$	$R^c$	計
$T$	20	20	40
$T^c$	16	24	40
計	36	44	80

ところが、ある医療関係者は、性別によって結果が異なるかも知れないと、新たに性別を考慮してデータを見たところ、左の表のような結果を得た。

このとき、男性 ( $M$ ) の場合、 $T$  を与えたときの治癒率は  $18/30 = 0.6$ 、与えないときのそれは  $7/10 = 0.7$  であり、女性 ( $F$ ) では  $T$  を与えたときの治癒率は  $2/10 = 0.2$ 、与えないときのそれは  $9/30 = 0.3$  となる。つまり、男性であれ女性であれ、 $T$  の治療法を受けさせないほうがよいという結果になる。

	$R_M$	$R_M^c$	$M$ 小計	$R_F$	$R_F^c$	$F$ 小計	計
$T$	18	12	30	2	8	10	40
$T^c$	7	3	10	9	21	30	40
計	25	15	40	11	29	40	80

上記の二つの例は統計をとる際、対象となる母集団の選択によっては正反対の結果になってしまうことがある、即ち、その選択の重要性を意味している。