

第2章：2次元データと単回帰分析

散布図と単回帰分析

以下は、無料通話アプリを使用したオンライン英会話・語学サイトにおける、顧客40人のレッスン受講回数データです。

分析担当者は、海外旅行インターネットメディア「TABINAVI」への広告出稿を検討するにあたり、以下のような仮説を立てました。

- 顧客のこれまでの海外旅行回数が多いほど、当社でのレッスン受講回数が増えるのではないか。

これを検証する手法として、単回帰分析を行う。以下の問に対する答えを検討してみましょう。

id	海外旅行回数	総受講回数
1	1	27
2	7	133
3	4	89
4	3	80
5	4	91
6	2	59
7	1	45
8	7	137
9	0	26
10	1	38
11	3	83
12	1	28
13	1	20
14	5	94
15	7	141
16	0	15
17	2	60
18	6	110
19	0	8
20	8	138
21	2	37
22	10	192
23	7	158
24	3	86
25	2	64
26	1	36
27	1	50
28	1	44

id	海外旅行回数	総受講回数
29	6	130
30	1	26
31	1	20
32	4	104
33	2	56
34	3	77
35	9	179
36	0	43
37	2	72
38	4	68
39	6	116
40	2	48

- Q1. 横軸（X軸）に「海外旅行回数」、縦軸に「総受講回数」をとり、散布図を描画してください。
- Q2. 2つのデータ列の相関係数を計算し、このまま単回帰分析を進められそうかを検討してください。
- Q3. 先ほど描いた散布図に、近似直線（線形近似）を引き、その直線の式を表示してください。

「TABINAVI」への広告出稿効果をシミュレーションするため、上記の単回帰モデルを用いて、さらに分析を進めます。

「TABINAVI」から取り寄せた媒体資料には、以下のような記述がありました。

- 「TABINAVI」ユーザの平均海外旅行回数：5.2回
 - 出典 2017年9月「TABINAVI」利用者調査

また、当社のレッスン受講単価（1コマ25分あたりの平均売上）は、500円とします。

以下の問に対する答えを検討してみましょう。

- Q4. 「TABINAVI」から獲得するユーザ1名の、当社レッスン総受講回数を予測してください。
- Q5. その回数にレッスン受講単価をかけ算することで、「TABINAVI」獲得ユーザ1人あたりの平均予測売上高を計算してください。
- Q6. 「TABINAVI」から新規ユーザを500名獲得出来た場合、当社の総売上はいくら増加すると予測できますか。

次の各問は、応用問題です。統計学とは直接関係がありませんので、スキップすることが出来ます。

- Q7. 「TABINAVI」から新規ユーザを500名獲得出来た場合、当社の変動費率を20%とすると、このプロモーションで赤字を出さない広告費の上限（広告の損益分岐点）はいくらですか。

(ここでは、追加の売上高から追加の変動費を引いた限界利益がマイナスであることを「赤字」と呼ぶこととします。)

- Q8. その上限（損益分岐点で広告を出稿した場合）での新規ユーザ獲得単価はいくらとなりますか。