

データに基づく推定

広島大学 AI・データイノベーション教育研究センター
稲垣知宏

目標

データとデータの間にある関係を仮定し、求めたい量を推定できるようにする

この授業で紹介すること

- データ間の関係
- データ間の関係から値を推定する

キーワード

データを起点としたものの見方、正負の相関

こんなことはありませんか？

県内のお好み焼き屋さんの数、
郵便局の数、スポーツ施設の数、
散髪屋さんの数が、それぞれの
程度あるか尋ねられたらどう
しますか？

- 知っていそうな人に聞く
- 地図を眺めて確認する
- データベースを探す



散髪屋さんの数を推定する

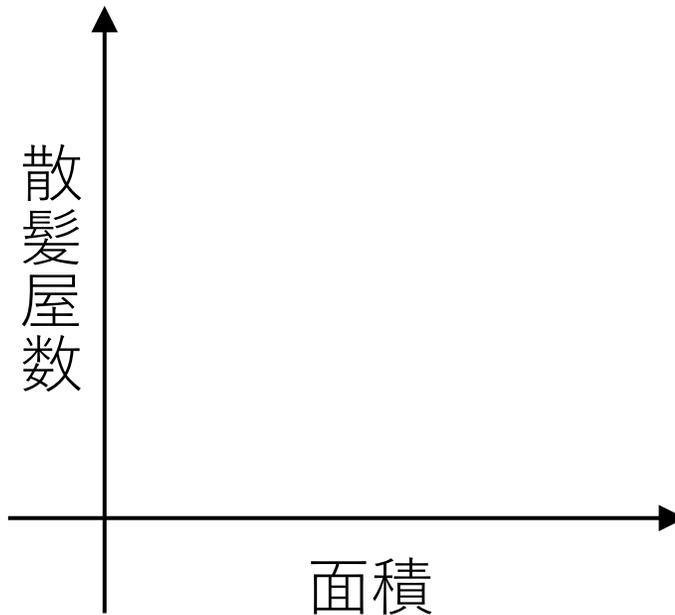
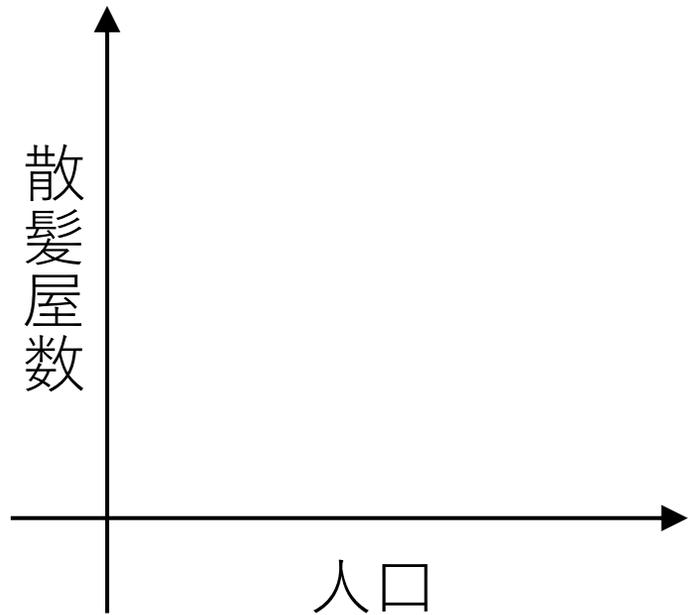
知っている、あるいは簡単に調べられるデータから、散髪屋さんの数を推定します。

簡単に調べられるデータとして、人口と面積からスタートしてみます。



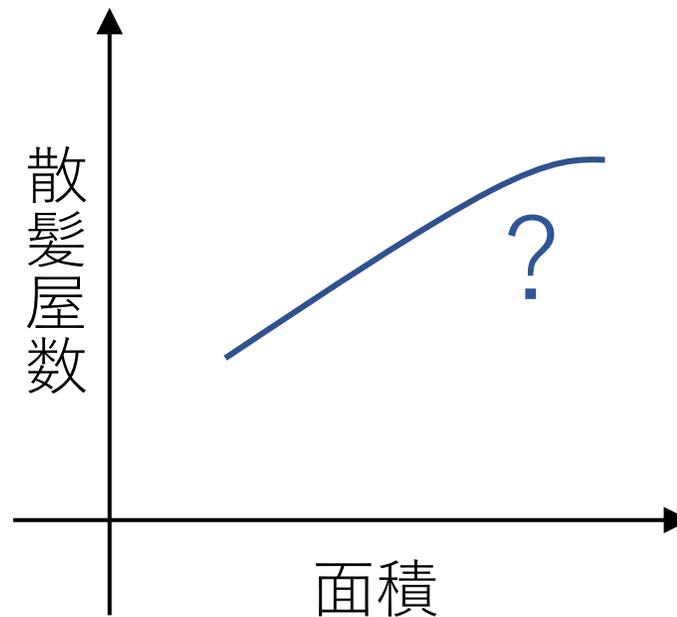
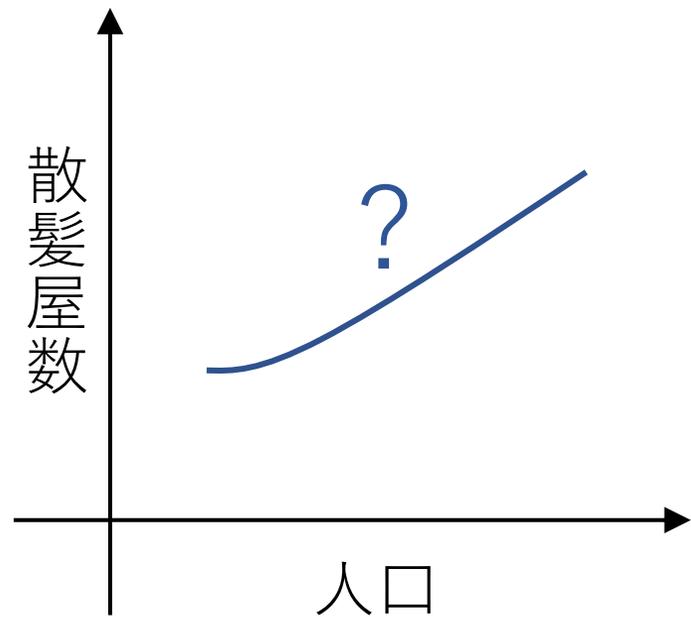
どのような関係？

人口、面積は散髪屋さんの数とどのような関係にあるでしょう。
以下の様なグラフを作成して、線を書き入れてみましょう。



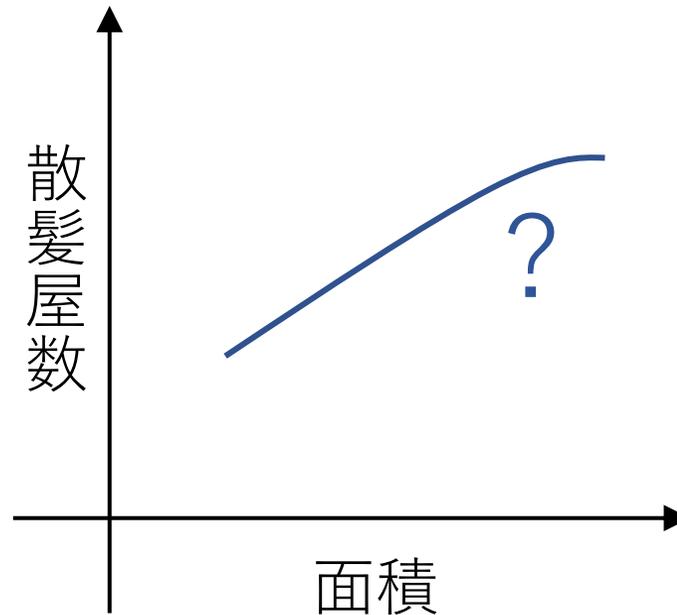
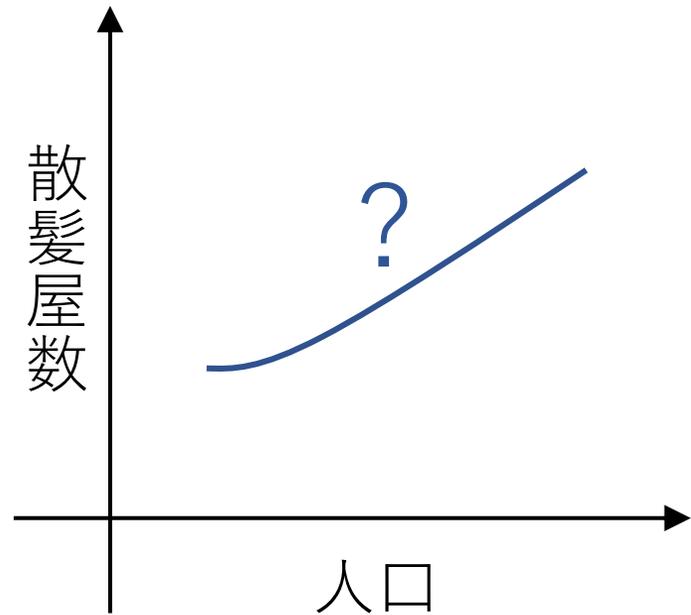
どのような関係？

人口、面積は散髪屋さんの数とどのような関係にあるでしょう。
以下の様なグラフを作成して、線を書き入れてみましょう。



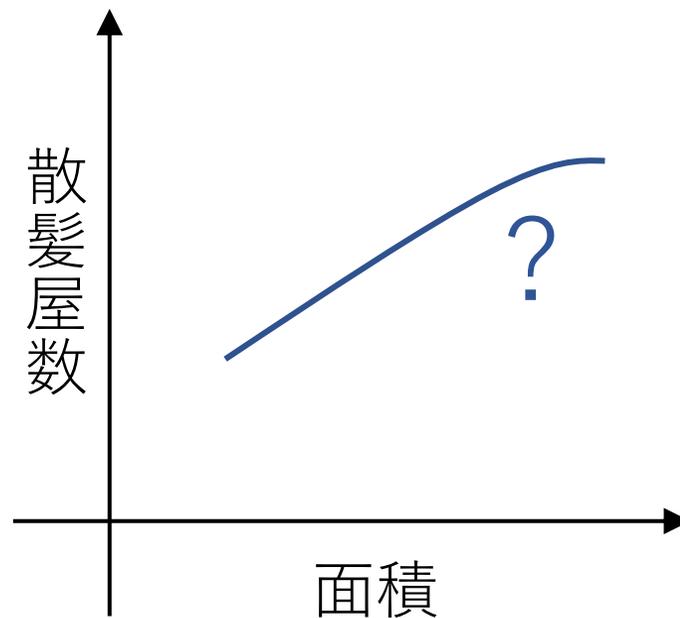
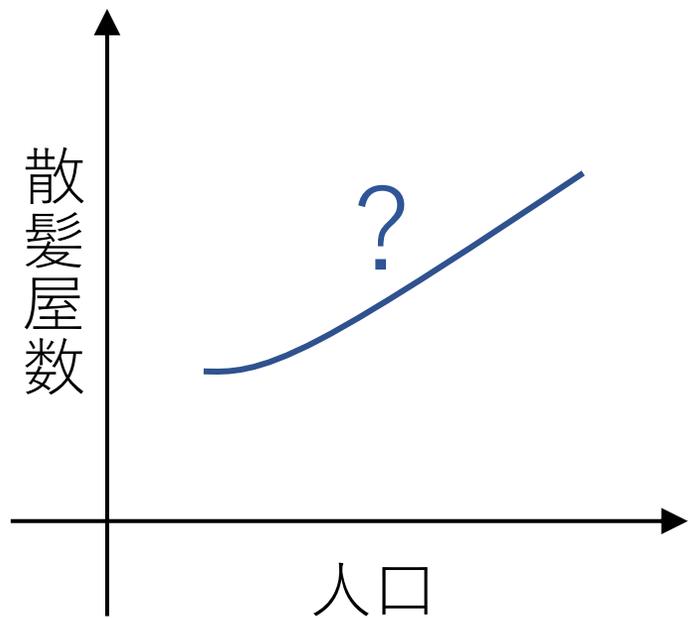
どのような関係があるのでしょうか？

人口、面積が増えると散髪屋さんの数も増えると考えたでしょうか。ある量が増えると、別の量も増える（減る）とき、それらの間には正（負）の相関があると言います。



どのような関係があるのでしょうか？

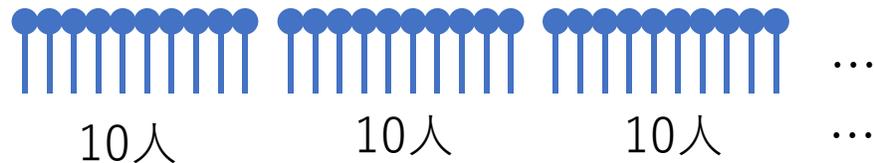
では、書き込んだグラフの傾きはどの程度になるでしょう？



傾きの推定

皆さんは、どの程度の頻度で散髪屋さんに行きますか？

散髪屋さんがやっていくためには、最低でも1日当たり10人程度のお客さんが来てくれる必要があるとします。



▶ 月当たりになると300人程度のお客さんが必要になります。

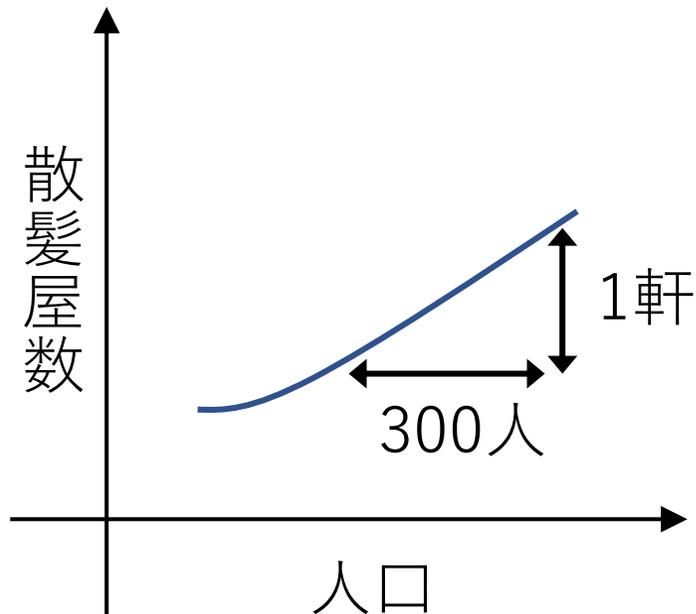
散髪屋さんに行く頻度は人それぞれで季節にもよると思いますが、おおよそ月に1回程度としてみます。

▶ 常連さんが300人いれば、月に300人来ると期待できます。

傾きの推定

1軒の散髪屋さんがやっていくには最低でも300人程度のお客さんが必要と推定しました。

▶ 300人当たり1軒程度の散髪屋さんがあってよい。



$$(\text{散髪屋数}) \sim \frac{1}{300} \times (\text{人口})$$

例題

考えた式

$$(\text{散髪屋数}) \sim \frac{1}{300} \times (\text{人口})$$

を用いて、国内の散髪屋さんの数を推定しなさい。

解説

散髪屋さんに通う人が国内に1億人いるとします。

これを推定した式に代入すると、

$$(\text{散髪屋数}) \sim \frac{1}{300} \times 100000000 \sim 33\text{万}$$

となります。

厚生労働省の衛生行政報告例：令和3年度衛生行政報告例の概況によると、理容所、美容所の施設数はそれぞれ、114,403と264,223（重複開設266）となっています。

施設数のデータに基づく推定

先ほどの例題では、散髪屋さんのおおよその数を推定し、実際の数字と大きく離れていないことを確認しました。

実際の散髪屋さんの数は約38万件ということですから、

$$(\text{散髪屋数}) \sim \frac{1}{300} \times 10000000 \sim 33\text{万}$$

の1億人が少なすぎるか分母の300が大きすぎるようです。

散髪屋さんに通う人が国内に1億人よりも多いというのは考え難く、分母の300が大きすぎると推定できそうです。実際の散髪屋さんの数からは、分母の値は264程度になります。

傾きを再度推定する

傾きの推定では、

- 最低でも1日当たり10人程度が来てくれる必要がある
- 散髪屋さんに行く頻度、おおよそ月に1回程度

の2つを仮定して1/300という傾きを求めました。これが1/264程度ということは、単純に考えると

- 1日当たりに来るお客の平均値は10人よりも少ない
- 散髪屋に行く頻度の平均値は月に1回程度よりも多い

となりますが、1日当たりのお客の平均値を9人、あるいは、週1日の休日を除いて平均10人にする等で、値を合わせられます。

問題

1. 流行っていると思う、スポーツ、もしくはゲームを1つ挙げなさい。
2. 1で挙げたスポーツ、あるいはゲームをしている人が県内にどの程度いそうか推定しなさい。

