

(.././index.html)

最小二乗法の意味と計算方法 - 回帰直線の求め方

Sci-pursuit (.././index.html) > 統計 (.././index.html#statistics)

> 最小二乗法の意味と計算方法 - 回帰直線の求め方

最小二乗法（または、最小自乗法）とは、誤差を伴う測定値の処理において、その**誤差の二乗の和を最小にすることで、最も確からしい関係式を求める**方法です。

最小二乗法の意味

最小二乗法とは、モデル関数を $f(x)$ とするとき、

$$\sum_{i=1}^n \{y_i - f(x)\}^2$$

が最小となるように $f(x)$ を求めることである。

モデル関数を 1 次関数とするとき、これを直線回帰といい、次のように式を求めることができます。

最小二乗法による回帰係数の計算方法

回帰直線を $y = ax + b$ とするとき

$$\begin{aligned} a &= \frac{s_{xy}}{s_x^2} \\ &= \frac{\sum_{n=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{n=1}^n (x_i - \bar{x})^2} \\ b &= \bar{y} - a\bar{x} \end{aligned}$$

ここで、

s_{xy} は x と y の共分散 (covariance.html)

s_x^2 は x の分散 (variance.html)

n は 2 変数データ (x, y) の総数

x_i と y_i は個々の数値

\bar{x} と \bar{y} はそれぞれの平均値

を表します。

このページの続きでは、直線回帰の場合を例に最小二乗法の意味と計算方法を、図を用いながら分かりやすく説明しています。



もくじ

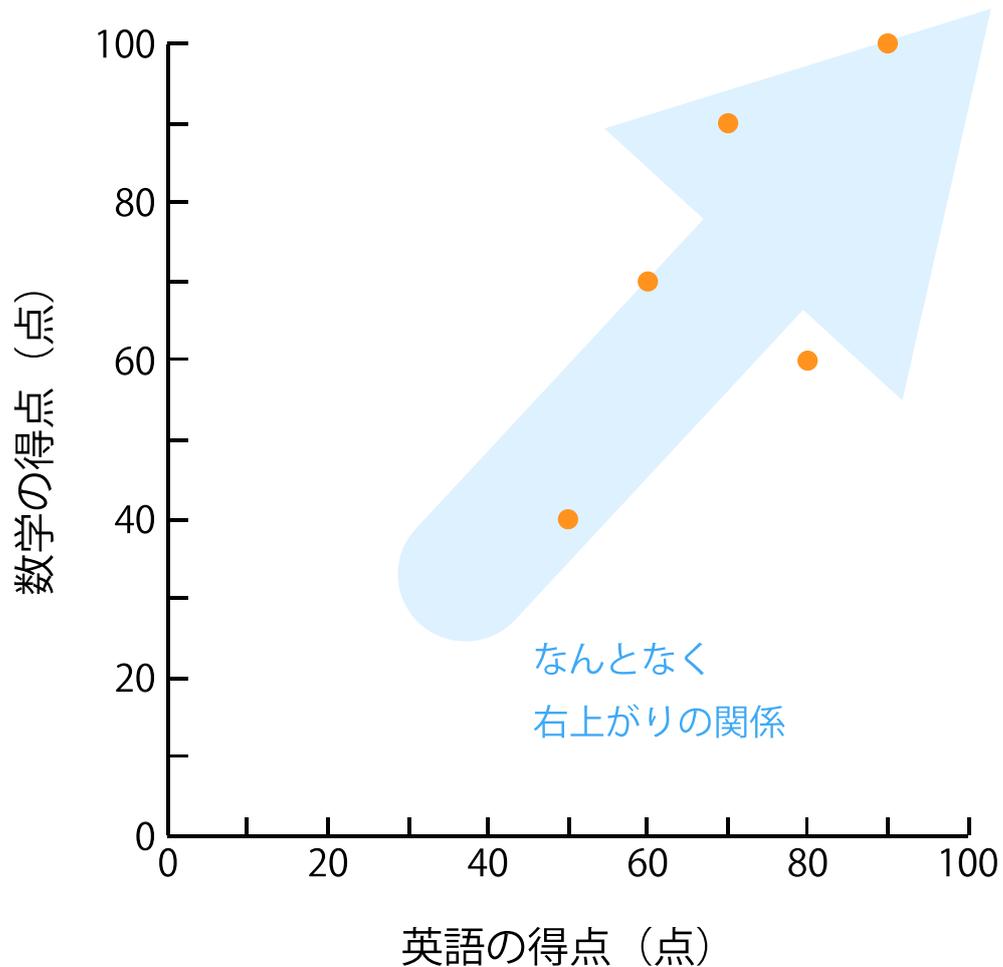
- 最小二乗法とは
- 最小二乗法による回帰直線の求め方

最小二乗法とは

最小二乗法（または、最小自乗法）とは、誤差を伴う測定値の処理において、その**誤差の二乗の和を最小にすることで、最も確からしい関係式を求める**方法です。

ここでは、最小二乗法によって回帰直線（1 次関数）を求める場合を例にとって、最小二乗法の説明をします。

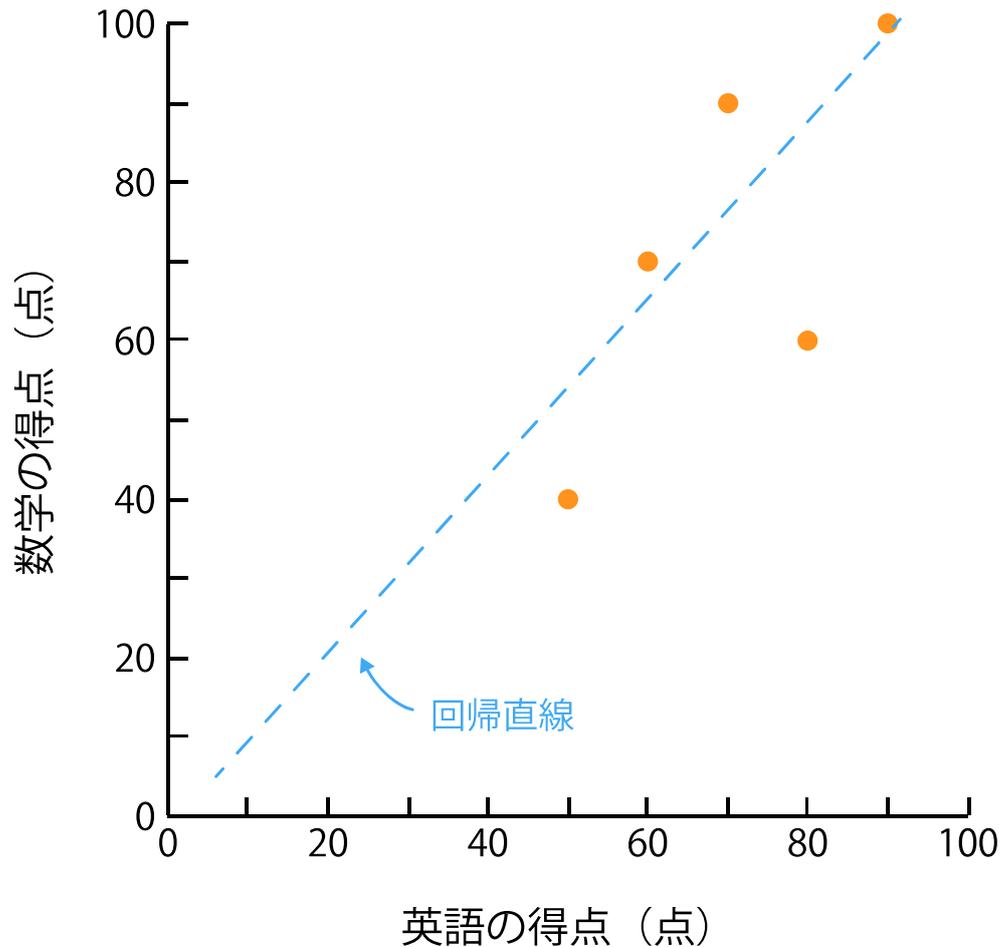
2 変数のデータの間に、次の散布図に示すような関係があったとします。例えば、変数として「テストにおける英語の得点と数学の得点」を考えてみましょう。その関係を、横軸に英語の得点、縦軸に数学の得点を取ってプロットしました。



英語と数学の得点の散布図 (右上がりの傾向がある)

散布図を見ると、プロット (点) はなんとなく右上がりになっていますね？つまり、英語の得点が高い人は数学の得点も高い、という関係が言えそうです。

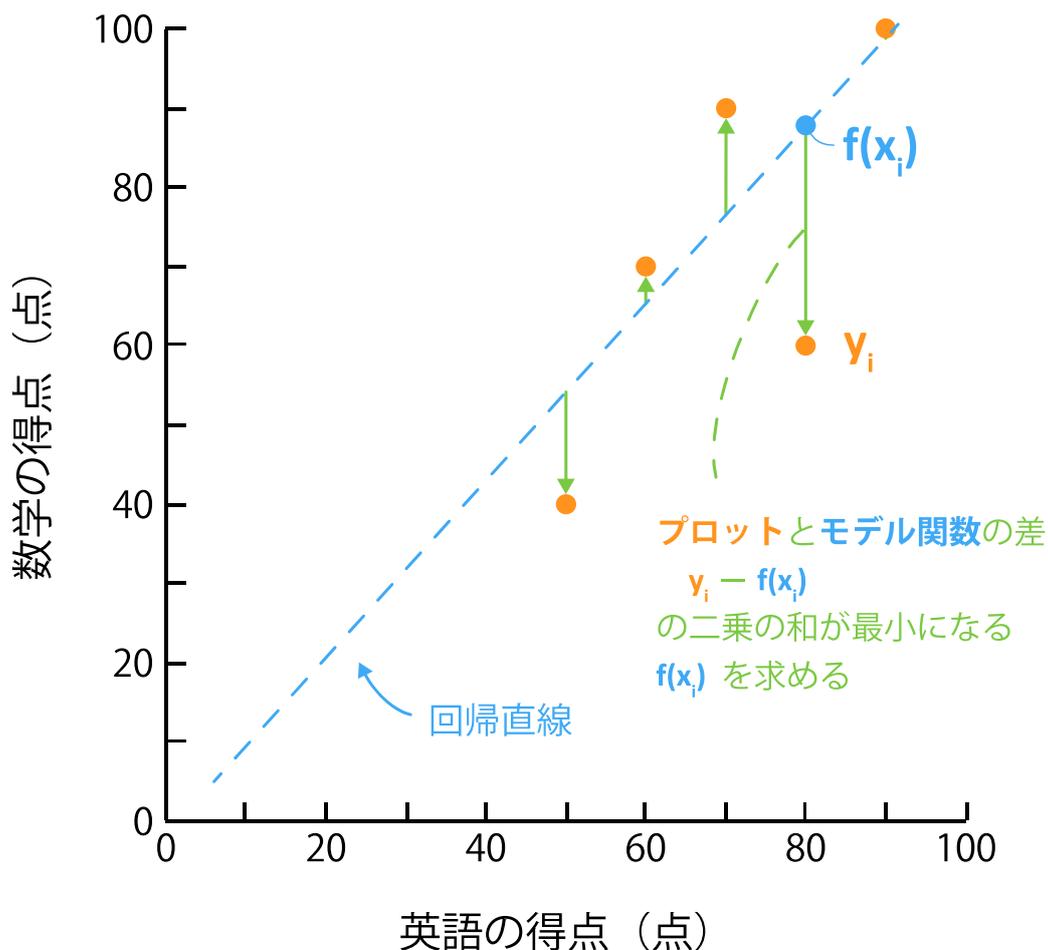
このような関係にもっともふさわしい直線を引くと、次のようになります。このような直線を引くことを、直線回帰といいます。



最小二乗法による回帰直線を書き加えた図

さて、このような右上がりの直線は「なんとなく」引くこともできますが、最小二乗法を使うことで、これを数学的な根拠を基に引くことができます。最小二乗法では、プロットの y 座標 (y_i) と、回帰直線上の y 座標 ($f(x_i)$) の差 (= 残差) の二乗 ($\{y_i - f(x_i)\}^2$) の和が最小になる関数 $f(x)$ を求めます。

つまり、下の図に示した緑色の矢印の長さの二乗の和が最小になる直線を求めます。



最小二乗法の原理

これを式で表すと、次のようになります。

最小二乗法の意味

最小二乗法とは、モデル関数を $f(x)$ とするとき、

$$\sum_{i=1}^n \{y_i - f(x)\}^2$$

が最小となるような $f(x)$ を求めることである。

最小二乗法を使って直線回帰をすると、このような 2 変数のデータに対して、もっともふさわしい唯一の直線を求めることができます。

続いては、最小二乗法を使って回帰直線を求めてみましょう。

Google Chrome - Chrome をダウンロード

Chrome はどの端末でも使える、高速で安全な無料のブラウザです。 google.co.jp



最小二乗法による回帰直線の求め方

最小二乗法で回帰直線を求めるには、次の計算を行います。

最小二乗法による回帰係数の計算方法

回帰直線を $y = ax + b$ とするとき

$$a = \frac{s_{xy}}{s_x^2}$$

$$= \frac{\sum_{n=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{n=1}^n (x_i - \bar{x})^2}$$

$$b = \bar{y} - a\bar{x}$$

ここで、

s_{xy} は x と y の共分散 (covariance.html)

s_x^2 は x の分散 (variance.html)

n は 2 変数データ (x, y) の総数

x_i と y_i は個々の数値

\bar{x} と \bar{y} はそれぞれの平均値

を表します。

この式にいきなり代入してもいいのですが、この後の例題では、次の 6 つのステップに分けて順番に計算していきます。

最小二乗法による回帰係数の求め方

1. それぞれの変数の平均値を求める
2. それぞれの変数の偏差 (deviation.html) (数値 - 平均値) を求める
3. 変数 x の分散 (variance.html) (偏差 (deviation.html) の二乗平均) を求める
4. 共分散 (covariance.html) (偏差 (deviation.html) の積の平均) を求める
5. 共分散 (covariance.html) を変数 x の分散で割って回帰直線の傾きを得る
6. 2 つの変数の平均値と傾きから、回帰直線の y 切片を得る

それでは、例題を一緒に解いて確認してみましょう。

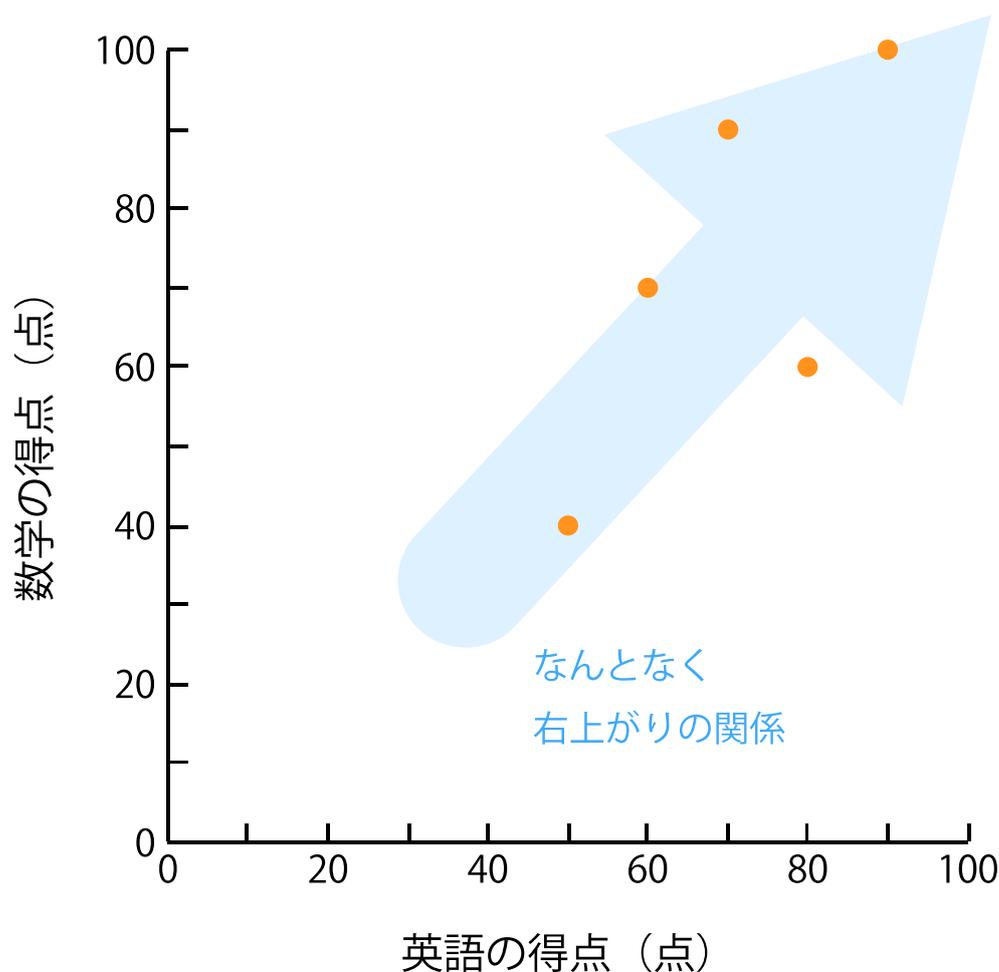
次に示した英語と数学の得点データの回帰直線を求めよ。

英語と数学の得点データ

	英語 (x)	数学 (y)
A さん	50 ($= x_1$)	40 ($= y_1$)
B さん	60 ($= x_2$)	70 ($= y_2$)
C さん	70 ($= x_3$)	90 ($= y_3$)

	英語 (x)	数学 (y)
D さん	80 ($= x_4$)	60 ($= y_4$)
E さん	90 ($= x_5$)	100 ($= y_5$)

はじめに、このデータの散布図を次に示します。



英語と数学の得点の散布図 (右上がりの傾向がある)

このプロットに最もふさわしい直線を最小二乗法によって求めましょう。

回帰係数を求めるため、平均値 → 偏差 (deviation.html) → 分散 (variance.html) → 共分散 (covariance.html) → 傾き → y 切片 の順で一つずつ求めていきます。

ステップ 1 : 平均値を求める

最初のステップでは、それぞれのデータの平均値 (\bar{x} と \bar{y}) を求めます。それぞれ計算すると、次のようになります。

$$\bar{x} = \frac{50 + 60 + 70 + 80 + 90}{5} = 70$$

$$\bar{y} = \frac{40 + 70 + 90 + 60 + 100}{5} = 72$$

よって、英語の平均点 (\bar{x}) は 70 点、数学の平均点 (\bar{y}) は 72 点と求まりました。

2018年03月03日(土) 19:00

V・ファーレン長崎

VS

サガン鳥栖

1ヵ月無料お試し

ステップ 2 : 偏差を求める

2 番目のステップでは、各データの偏差 (deviation.html)を計算します。偏差 (deviation.html) とは、各データの数値から平均値を引いた差のことです。すなわち、変数 x に対しては $x_i - \bar{x}$ 、変数 y に対しては $y_i - \bar{y}$ を計算します。偏差について詳しくは「偏差の意味と求め方 (deviation.html)」をご覧ください。

例えば、A さんの英語の偏差 (deviation.html)は

$$(A \text{ さんの英語の得点 } x_1) - (\text{英語の平均値 } \bar{x}) = 50 - 70 = -20 \text{ (点) と求まります。}$$

A さんの数学の偏差 (deviation.html)は

$$(A \text{ さんの数学の得点 } y_1) - (\text{数学の平均値 } \bar{y}) = 40 - 72 = -32 \text{ (点) となります。}$$

このようにして、B さんから E さんまでの偏差 (deviation.html)を、英語と数学についてそれぞれ計算したものが次の表です。

英語の得点データと平均値、偏差

	英語の点数 (x_i)	偏差 (deviation.html) ($x_i - \bar{x}$)
A さん	50 (= x_1)	-20 (= $x_1 - \bar{x}$)
B さん	60 (= x_2)	-10 (= $x_2 - \bar{x}$)
C さん	70 (= x_3)	0 (= $x_3 - \bar{x}$)
D さん	80 (= x_4)	10 (= $x_4 - \bar{x}$)
E さん	90 (= x_5)	20 (= $x_5 - \bar{x}$)
平均値	70 (= \bar{x})	—

数学の得点データと平均値、偏差

	数学の点数 (y_i)	偏差 (deviation.html) ($y_i - \bar{y}$)
A さん	40 (= y_1)	-32 (= $y_1 - \bar{y}$)
B さん	70 (= y_2)	-2 (= $y_2 - \bar{y}$)
C さん	90 (= y_3)	18 (= $y_3 - \bar{y}$)
D さん	60 (= y_4)	-12 (= $y_4 - \bar{y}$)
E さん	100 (= y_5)	28 (= $y_5 - \bar{y}$)
平均値	72 (= \bar{y})	—

ステップ 3 : 変数 x の分散を求める

3 番目のステップでは、変数 x の分散 (variance.html) を求めます。分散 (variance.html) とは、偏差 (deviation.html) の二乗の平均のことです。分散 (variance.html) について詳しくは「分散の意味と求め方 (variance.html)」をご覧ください。

前のステップで求めた偏差 (deviation.html) を用いて、英語の得点の分散 (variance.html) s_x^2 は次のように計算できます。

$$s_x^2 = \frac{(-20)^2 + (-10)^2 + 0^2 + 10^2 + 20^2}{5} = 200$$

よって、英語の得点の分散 (variance.html) s_x^2 は 200 (単位 : 点²) と求まりました。

ステップ 4 : 共分散を求める

5 番目のステップでは、共分散 (covariance.html) を求めます。共分散 (covariance.html) とは、偏差 (deviation.html) の積 $(x_i - \bar{x})(y_i - \bar{y})$ の平均値です。共分散 (covariance.html) について詳しくは「共分散の意味と求め方 (covariance.html)」をご覧ください。

偏差 (deviation.html) は 2 番目のステップで既に求めました。この計算結果およびその積を下の表に示します。

なお、偏差 (deviation.html) の積とは、例えば A さんについて、A さんの数学の偏差 (deviation.html) (-20 点) と英語の偏差 (deviation.html) (-32 点) を掛け合わせたものです。この計算結果は、 $(-20) \times (-32) = 640$ (単位 : 点²) になります。

同様に B さんから E さんまで偏差 (deviation.html) の積を計算した結果を次の表に載せています。

英語と数学の偏差とその積

	英語の偏差	数学の偏差	偏差の積
A さん	-20	-32	640
B さん	-10	-2	20
C さん	0	18	0
D さん	10	-12	-120
E さん	20	28	560

偏差 (deviation.html) の積の平均を取ったものが、共分散 (covariance.html) です。上の表より、共分散 (covariance.html) s_{xy} は次のように計算できます。

$$\begin{aligned} s_{xy} &= \frac{640 + 20 + 0 + (-120) + 560}{5} \\ &= 220 \end{aligned}$$

よって共分散 (covariance.html) (s_{xy}) は 220 (単位: 点²) と求まりました。

決算 SALE
2018年3月31日申込分まで!!



ステップ 5 : 傾きを求める

ステップ 3 で求めた英語の得点の分散 (variance.html) s_x と、ステップ 4 で求めた共分散 (covariance.html) s_{xy} は以下の通りでした。

$$\begin{aligned} s_{xy} &= 220 \text{ 点}^2 \\ s_x^2 &= 200 \text{ 点}^2 \end{aligned}$$

よって、最小二乗法によって回帰係数を求める公式より、回帰直線の傾き a は次のように求まります。

$$\begin{aligned} a &= \frac{s_{xy}}{s_x^2} \\ &= \frac{220}{200} \\ &= 1.1 \end{aligned}$$

したがって、回帰直線の傾き a は 1.1 と求まりました

ステップ 6 : y 切片を求める

最後に、回帰直線の y 切片 b を求めます。ステップ 1 で求めた平均値 \bar{x} , \bar{y} と、ステップ 5 で求めた傾き a を、回帰直線を求める公式に代入します。

$$\begin{aligned} b &= \bar{y} - a\bar{x} \\ &= 72 - 1.1 \times 70 \\ &= -5.0 \end{aligned}$$

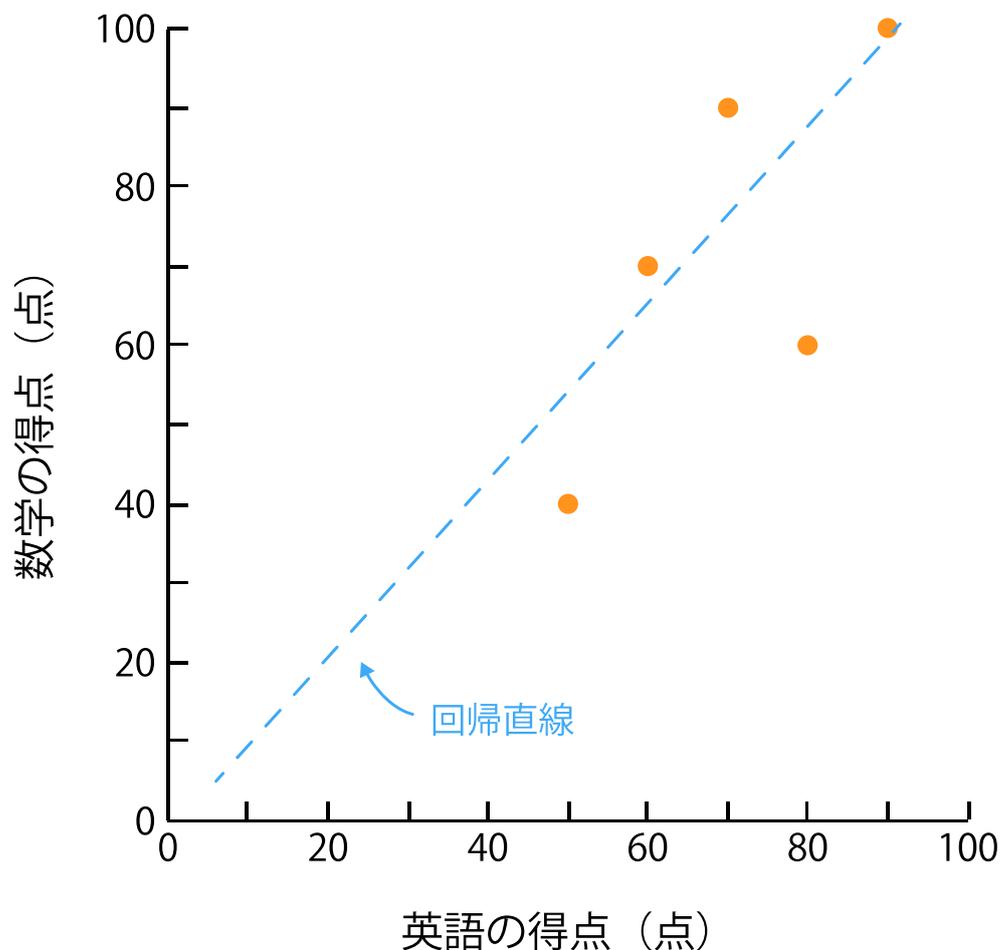
よって、回帰直線の y 切片 b は -5.0 (単位 : 点) と求まりました。

最後に、傾きと切片をまとめて書くと、次のようになります。

$$y = 1.1x - 5.0$$

これで最小二乗法に基づく回帰直線を求めることができました。

散布図に、いま求めた回帰直線を書き加えると、次の図のようになります。



最小二乗法による回帰直線を書き加えた散布図

人工知能特化オンライン塾

8週間の集中特訓で人工知能のプロフェッショナルになろう。オーダーメイドのカリキュラムを提供 aidemy-premium.net



- 化学実験器具一覧
(../chem/labware/index.html)
- 浸透圧はなぜ生じるのか？
(../chem/osmotic_pressure-1.html)
- 石灰水の作り方 (../chem/limewater-1.html)
- 微分とは何か？ - 微分のイメージ
(../differential-1.html)
- 浸透圧の計算方法
(../chem/osmotic_pressure-2.html)
- 石灰水の反応のしくみ
(../chem/limewater-2.html)

Sci-pursuit.com (../index.html) 2017 All rights reserved.

[トップページ \(../index.html\)](http://../index.html) | [プライバシーポリシー \(../info/privacy_policy.html\)](http://../info/privacy_policy.html) | [免責事項 \(../info/disclaimer.html\)](http://../info/disclaimer.html)