2015年度春学期　統計学ー数学の補足説明　第3回

回帰分析に関連して

関数とパラメータ

関数とは，「ある数を入れると，何か別の数が出てくる」計算のことです。これを昔は，「ひみつの箱」の中にある数を入れたら，箱から別の数が出てくるという意味で「函数」と書きました。「函」は，「函館」を「はこだて」と読むように，「箱」と同じ意味です。

入れる数を変数 x で表し，出てくる数が変数 y であるような関数の名前が f であることを，$y = f(x)$ と書きます。関数の名前に f をよく用いるのは，関数にあたる英語 "function" の頭文字をとっているためです。他に，アルファベットでその続きである g, h もよく用います。

関数は計算ですから，その計算が簡単な式で表されるものもよくあります。もっとも簡単な関数は，変数 x になんらかの定数をかけて，さらに何らかの定数を足すという計算で表される「一次関数」です。これを式で表すと，定数 a と定数 b を使って $y = a + bx$ となります。

定数 a, b がどんな数字であっても，この関数が一次関数であることには変わりはありません。定数 a, b は，一次関数が $y = 2x + 3$ なのかあるいは $y = -4x - 2$ なのか，つまり「どんな一次関数」かを決める数字です。このような定数を**パラメータ**といいます。

関数とグラフ

関数 $y = f(x)$ がどんな式で表されるかがわかれば，x がいくらのときに，その x を入れると y がいくらかになるかを計算することができます。そこで，座標平面の横軸で x を表し，縦軸を y で表すと，「x がいくらのとき y はいくら」という関係は，座標平面での線（グラフ）で表されます。グラフの形は，その関数が行う計算の種類によって特徴があります。たとえば，一次関数なら直線になりますし，$y = ax^2 + bx + c$ のように x^2 を含む式で表される関数（二次関数）ならば，放物線という曲線になります。

また，「変数 x と変数 y を入れると，変数 z が出てくるような関数」を考えて，$z = f(x, y)$ と表すこともできます。「回帰分析」のところで出てくる「最小二乗法」では，このタイプの関数を扱っています。このような関数を表すグラフは，x, y, z の3つの軸でできた座標「立体」の中にある「曲面」になります。

微分

関数 $y = f(x)$ について，x がどれだけ変化したら，それに応じて y がどれだけ変化するか，という問題を考えます。グラフを使ってみてみましょう。

図1で，x が定数 x_1 から定数 x_2 まで変化するとき，y は定数 y_1 から定数 y_2 まで変化しています。このとき，(y の変化量）の（x の変化量) に対する割合を**変化率**といいます。この場合だと，変化率は $\dfrac{y_2 - y_1}{x_2 - x_1}$ となります。いま求めた変化率は，x が x_1 から x_2 まで，幅 $(x_2 - x_1)$ だけ変化するときの変化率です。グラフでいうと，このグラフが坂道で，その上を x_1 から x_2 まで歩くと考えるとき，その間がどれくらいキツい斜面であるかを表しています。変化率が $+$ なら登り坂，$-$ なら下り坂です。

いま求めた変化率は，x_1 から x_2 までを歩く間での変化率です。歩いている間に，斜面のキツさは変化するかもしれません。実際，図1の例では，x_1 から x_2 まで歩くと，斜面はだんだんキツくなっています。では，x_1 に立って止まっているときの「足元」の斜面のキツさはどうやって求めればよいでしょ

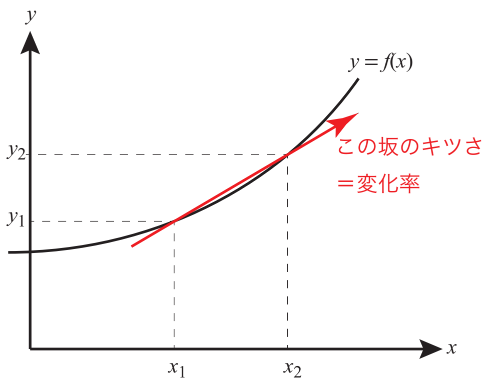

図 1: 変化率

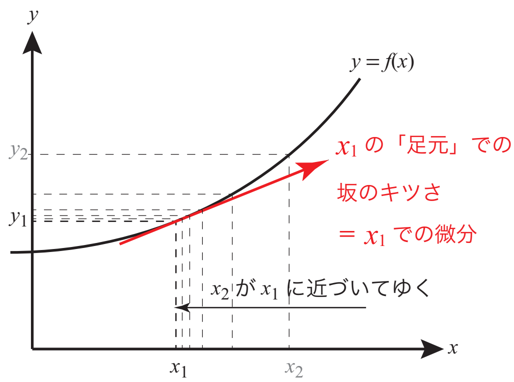

図 2: 微分

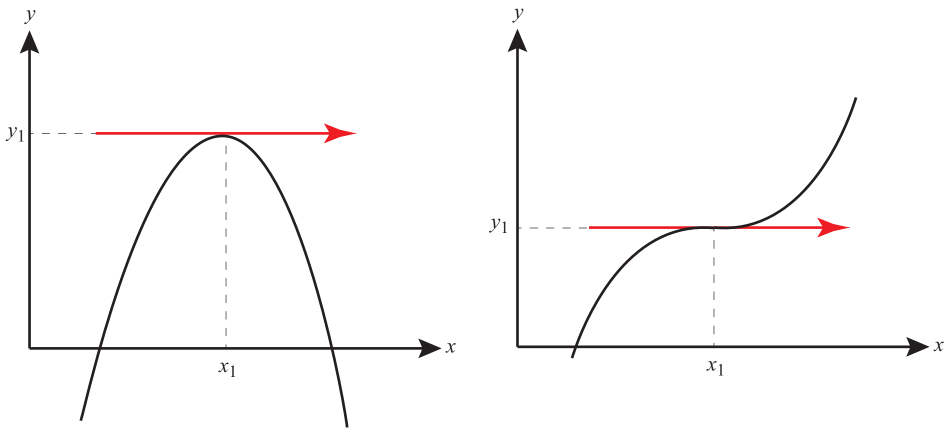

図 3: 微分が 0 のとき

うか。x_1 に立ち止まっていることから，変化率を求める $\dfrac{y_2 - y_1}{x_2 - x_1}$ という式で，「x_1 から x_1 まで歩く」として，つまり x_2 に x_1 を代入して計算すると，この分数は $\dfrac{y_1 - y_1}{x_1 - x_1}$ になり，分母が 0 になるので計算できません。

そこで，x_2 が x_1 に「近づいてくる」という状況を考えましょう。x_2 が x_1 に近づいて，幅 $(x_2 - x_1)$ が狭くなっても，変化率は同じように上の分数で求めることができます。そして，図 2 のように x_2 が x_1 と重なってしまった場合でも，「幅 $(x_2 - x_1)$ がどんどん狭くなって，0 にはならないが 0 に限りなく近づいていく」と考えることで，「x_1 に立っているときの，足元での変化率」を求めることができます。これを，「関数 $y = f(x)$ の，$x = x_1$ での**微分**」といいます。

ところで，x がある値のときに微分が 0，つまり「足元での斜面のキツさが 0」とはどういうことでしょう。このときは，足元の斜面が登りでも下りでもないわけですから，図 3 のように，足元がグラフの頂点や踊り場になっていることになります。「回帰分析」での「最小二乗法」の説明では，この性質を使っています。

昇べき順と降べき順

ところで，高校までに一次関数を習ったときに，一次関数の式を $y = ax + b$ と習った人も多いとおもいます。一方，統計学では $y = a + bx$ のような書き方がよく用いられます。二次関数を書くときも，前者のように $y = ax^2 + bx + c$ と書くやり方と，後者のように $y = a + bx + cx^2$ と書くやり方があります。前者のように，x の肩についている指数の大きいものから順に書く方法を**降べき順**，後者のように，指数の小さなものから順に書く方法を**昇べき順**といいます[1]。

統計学で昇べき順がよく用いられるのは，単純な式でまず考えて，それからいろいろな要素を付け加えてだんだん複雑なことを考えるというやり方をすることが多いので，複雑なことを後ろに書くほうが書きやすいためです。

[1]「べき」は漢字では「冪」です。