■■連載記事■

データをビジネスに活用する実践アナリティクス

<第 8 回>　変化を鋭く捉えるには

梶山 昌之
株式会社ワイハット

1. 統計的検定が必要な理由

販売業では製品のキャンペーンを行った結果，売上が向上したことを確認することが課題になります．製造業では生産性向上の施策の効果があったことを示す必要があるでしょう．医学では薬の安全性や有効性を確認する必要があります．

しかし，測定の対象とする量はバラツキを持つものですので，例えば，先月に比べて今月の生産性が高かったとしても，それは生産性向上施策とは無関係の偶然の結果かもしれないのです．このような場合に統計的検定の手法が必要になります．前回の連載では，技術研修による技術力向上の例で，この問題を考えました．

ここで，技術力は技術研修で行うテストの成績で測定するものとします．この場合注意すべき点は，技術研修の効果がまったくない場合でも，偶然の結果，テストで高い得点を取る場合があるということです．また，基本的な問題として，特定の点数を取れる実力になっているかを考える問題について考えました．

テストを無限に繰り返せばクラスの実力の分布が明らかになり，この問題に答えることができるでしょうが，それは不可能です．ここで，クラスの成績と言わずにクラスの実力と表現したのは，テストの成績は実力を反映するが，実力そのものではないという考えからです．この場合の実力とは，クラスの成績の母平均と同じです．

それでは，1 回のテストの結果で判断するにはどうしたらよいでしょうか．

このような問題は実務でよく見かけるものであり，一見簡単そうに見えるのですが，これを正しく解こうとすると，「平均の分布」「不偏推定量」「t 分布」の用語の意味を理解することが必要でした．

この内容は，やや難しいので，統計の学習もこの辺で挫折してしまう人が多いようです．しかし，理論的な背景については，数式の展開よりは，直観的に内容を理解できていれば十分です．t 検定の手順に従えば結果が得られますので，後は結果をどのように解釈するかという点が重要です．

2.　データ数と検出力

ここで，前回の復習も兼ねて，以下の問題について考えてみてください．

「技術研修でクラス A の得点の平均が 50 点になることを目標として研修を続けているとします．テストを行い 7 名分の得点が得られました．このクラスの真の実力は平均 50 点と言えるでしょうか．」

表 1　クラス A の得点

No	氏名	得点
1	A	48
2	B	54
3	C	58
4	D	47
5	E	58
6	F	56
7	G	57
標本平均		54.0
不偏分散		21.67

表 1 には不偏分散が計算されています．不偏分散を Excel で計算する場合は，VAR.S 関数を使用してください．実はこの問題は，前回が 8 名のデータであったのに対して，最後の 1 名を除いたデータです．この場合にも果たして同じ結果が得られるでしょうか．

この値から t 値を計算すると

$$t_o = \frac{\bar{x} - \mu_o}{\sqrt{s^2/n}} = \frac{54 - 50}{\sqrt{21.67/7}} = 2.273$$

となります．このように，標本データから計算され，検定のために用いる統計量は検定統計量と呼ばれます．このケースの場合の検定統計量 t_0 は標本平均が基準となる値（50 点）から隔たっている程度を表す量と理解してください．この値が，クラスの実力の変動を示すほど大きな値かどうかを調べましょう．

t 分布表（表 2）の両側確率 5%自由度 6 の値を読み取ると，2.447 になります．すなわち，検定統計量 t_0 は判定の基準となる t 値以上となりませんでした．

このことから，テストの結果は，母平均が 50 点の場合にも起こりうる結果であり，クラスの実力が向上しているとは言えないという結論になりました．

表 2　t 分布表（パーセント点）

自由度	両側確率 P			
df	0.1	0.05	0.02	0.01
1	6.314	12.706	31.821	63.657
2	2.920	4.303	6.965	9.925
3	2.353	3.182	4.541	5.841
4	2.132	2.776	3.747	4.604
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169
11	1.796	2.201	2.718	3.106
12	1.782	2.179	2.681	3.055
13	1.771	2.160	2.650	3.012
14	1.761	2.145	2.624	2.977
15	1.753	2.131	2.602	2.947
16	1.746	2.120	2.583	2.921
17	1.740	2.110	2.567	2.898
18	1.734	2.101	2.552	2.878
19	1.729	2.093	2.539	2.861
20	1.725	2.086	2.528	2.845

このケースの様にデータ数が少なくなると，実力が変化していてもそれを検出できない場合があります．

この場合の対立仮説は H_1: $\mu \neq 50$ です．対立仮説が正しい場合に，帰無仮説 H_0: $\mu = 50$ を棄却する確率を検出力と呼びます．検出力を高くする方法の一つが分析対象のデータ数を増やすことです．データ数が増えると小さな変化でも検出できるようになります．

このケースの場合「クラスの実力は向上していない」ことを積極的に証明できたわけではなく，得られたデータでは向上を示すことができないとう状況ですので，結果は「クラスの実力が変わった（向上した）とは言えない」と表現してください．

3.　両側検定と片側検定の使い分け

前節で解説した検定の手順は「両側検定」と呼ばれているものです．技術力が向上したかどうかを検定したいのですが，研修の影響は，技術力を向上させる効果がまったくないかまたは低下させる可能性もあることを想定しています．

技術的な知見により変化の方向がわかっている場合は別ですが，それが明示されていない場合には「両側検定」を用います．それでは，次の問題について考えてみてください．

「7 名の生徒のクラスの技術レベルは過去の実力テストで平均 50 点であった．ある技術研修は技術レベルを向上させる効果があることが知られており，この研修を行った後に実力テストを実施した．このクラスの実力は向上したといえるだろうか.」

テストの得点は表 1 と同じであるとします．

この問題の場合,「技術レベルを向上させる効果があることが知られており」と記述されている点がポイントです．研修により技術力が低下する可能性については考えなくてよいため，有意水準 5%で判定する場合でも，分布の片側（上側確率）が 5%になる点で判定すればよいことになります．この場合，帰無仮説は H_0: $\mu = 50$，対立仮説は H_1: $\mu > 50$ となります．

両側確率 P からパーセント点を求める t 分布表を使用する場合，両側で 10%の値を読めば，片側は 5%になります．この関係を利用して，片側確率のパーセント点を求めるために，両側確率の表を利用することができます．

t 分布表で両側確率 10%で自由度 6 の値を読み取ると，1.943 になります．その結果，データから計算された t 値の方が大きくなるため,「クラスの実力が向上したと言える」という結論になりま

す．このように片側検定を採用できれば，検定の検出力が高くなります．

両側検定を行うか，片側検定を適用可能かは統計の問題ではなく，分析対象に対する技術的知見の問題になります．

たとえば，ネジを製造する場合は，ネジの直径は大きすぎても小さすぎても不良品になりますので，両側検定を使います．

入社試験の成績が前年度に比べて変化があるかどうかを調べる場合は，向上と低下の両方の可能性がありますので両側検定を使用すべきでしょう．生産性向上施策の効果を検定する場合は，施策の逆効果の要素がないと判断できれば，片側検定が採用できます．

4.　二つの集団の比較

前節までの解説は母数（母平均）が任意の値（50点）と同じか，あるいはそれよりも大きいか，それとも小さいかどうかを検定するためのものでした．より一般的な問題は，実施前と実施後を比較する問題，または 2 つの集団を比較する問題でしょう．そこで，次は以下の問題について考えてみてください．

「技術研修のクラス A，B でテストを行った．これらのクラスの実力に差があると言えるだろうか.」

表 3　クラス A と B の得点

No	A	B
1	48	36
2	54	44
3	58	56
4	47	34
5	58	42
6	56	44
7	57	49
8		41
9		59
標本数	7	9
標本平均	54.0	45.0
偏差平方和	130.00	562.00
不偏分散	21.67	70.25

ここでは，分析対象のデータは以下の前提を満たすものとします．

① A と B の得点は正規分布に従う．
② A の分散と B の分散は等しい．

これらの前提の元に，A の平均と B の平均の差を判定するための検定統計量は，

$$t=\frac{(\bar{x}_A-\bar{x}_B)-(\mu_A-\mu_B)}{\sqrt{\left(\frac{1}{n_A}+\frac{1}{n_B}\right)}\sqrt{\frac{S_A+S_B}{n_A+n_B-2}}}$$

となります．ここで，S_A, S_B は偏差平方和で，Excel では DEVSQ 関数で求めることができます．各不偏分数は，この偏差平方和を自由度で割ったものです．

一見，難しい式に見えますが，分子の部分は標本平均と母平均の差で，分母は二つの集団の標準偏差を合成した標準偏差（合成標準偏差）です．従って、全体としては標準化の式になっていると理解してください．ここで，2 つのクラスの母平均に差がないとすれば，帰無仮説は H_0: $\mu_A=\mu_B$ です．計算された標本平均と不偏分散を使って統計検定量を求めると，

$$t_o=\frac{(54-45)}{\sqrt{\left(\frac{1}{7}+\frac{1}{9}\right)}\sqrt{\frac{130+562}{7+9-2}}}=2.540$$

となります．自由度 14 (=7+9-2) の両側確率 5% におけるパーセント点は 2.145 ですので，観測された t 値の方が大きくなっています．このことから，A と B のクラスの実力には差があり，A の方が B より 9 点高いと言えます．

実際はデータが完全な正規分布に従うとは考えられないのですが, t 検定ではデータが多少正規分布からはずれていても結論には大きな影響はないことが知られています．このことは t 検定の頑健性（ロバストネス）と呼ばれています．

また，2 つの集団が正規分布に従う場合でも，その分散は異なる場合がありますが，ここでは分散が等しいとして分析を行いました．分散の大きさが異なっているかどうかを知るためには，分散の検定を行う必要があります．また，分散が異なる場合には，その後の t 検定の方法も多少異なったものになります．

5.　対応のあるデータの検定

クラスBはクラスAより実力が低いことがわかりましたので，クラスBには追加の技術研修と再試験を行いました．

ここで次の問題について考えてみてください．

「クラス B の技術力向上を目的として技術研修を行った．この技術研修は技術レベルを向上させる効果があることが知られており，この研修を行った後に2回目の実力テストを実施した．このクラスの実力は向上したといえるだろうか.」

表4　クラスBの得点

出席番号	1回目	2回目	増加分
1	36	43	7
2	44	48	4
3	56	56	0
4	34	32	−2
5	42	40	−2
6	44	56	12
7	49	55	6
8	41	53	12
9	59	59	0
標本数	9	9	9
標本平均	45.0	49.1	4.1
偏差平方和	562.00	656.89	244.89
不偏分散	70.25	82.11	30.61

クラスAとクラスBとの比較と同じ様に，1回目の結果と2回目の結果を比較してみます．

統計検定量は0.999となり，両側確率10%で自由度16の値を読み取ると1.746になります．その結果「クラスの実力が向上したとは言えない」という結論になります．

しかし，一人ひとりのデータをよく見ますと，成績が上がっていない生徒もいますが，向上した生徒も多い様です．研修に効果がないならば，この増加分の平均は0になると考えられますが，平均4点増加しています．このケースの場合は，この増加分のデータを対象として t 検定を適用する方式が正解です．

クラスAとクラスBの比較の場合と異なる点は，1回目と2回目のデータは同じ生徒のデータですので，対応があるという点です．研修に効果がないならば，増加分の母平均は0になりますので，帰無仮説は H_0: $\mu=0$ となります．研修は技術レベルを低下させることがないと考えるなら，対立仮説は H_1: $\mu>0$ です．

増加分の9つのデータから t 値を計算すると2.229となります. この値はt分布表の片側確率5%（両側確率10%）自由度8の値1.860よりも大きいので,「このクラスの個人の実力は変わっていないとは言えない（向上した）」という結論になります．

データに対応がある場合は，個々のデータの変動分に着目するほうが，変化を検出しやすくなることがわかります．

今回は2つの集団の差を検定する方法について学びましたが，次回は，ある商品が好きかどうかを調査したアンケートを分析する方法について解説の予定ですのでご期待ください．

最後に本稿の理解に役立つ文献を参考文献（今泉忠ほか, 2012）として挙げます．

参考文献

今泉忠，田村義保，中西寛子，美添泰人 (2012). 日本統計学会公式認定 統計検定2級対応 統計学基礎．東京図書．