エクセルでの平均の差の検定について
Statistical Test of Means between Two Pupulations

(株) 竹内ハガネ商行　　竹内寿一郎

1．はじめに

従来、耳鼻咽喉科の先生方とともに、医学データの統計処理に携わってきたが、統計的仮説検定の問題のほとんどは、２つの母集団における母平均の違いの検定と言って良いくらいであった。この検定は大昔は大数に基づく正規分布の検定であり、1800 年初頭の t 分布の発見以来、分散が未知の場合は必ず t 検定が使われてきた。近年になって何でも正規分布、かつ等分散である仮定に疑問が生じて、分散が異なる場合、その和の近似自由度が、$Satterthwaite$ によって計算されるようになり、$Welch$ による検定が推奨されるに至ったのである。

エクセルにおける統計的検定法の分析ツールもそれなりに変化を遂げてきて、単純な正規検定、t 検定は削除され、最近では 2 標本に基づく母平均の検定法に限られ、分析ツールにおさめられている。

ここで、いくつかの事柄について注意を述べておく。データの分布は正規分布に従うと仮定して、さらに、

(1)　２つの母集団の分散 (標準偏差) が既知の場合もあるとして考慮しておく。分散は本来神様のみにしか分からない数値であるが、ここでは長年の経験から既知であるという立場に立つことにする。

(2)　ところで、2 つの母集団の間で、分散が等しいということの裏付けはあるのか。

(3)　もしも 2 つの母集団で分散の大きさが異なった場合、それぞれの分散はいくつと見積ればよいか。

(4)　分散が未知の場合、2 つの母集団の分散が等しいという仮定に則って標本から推定するために最適な推定値はどうするか。

(5)　分散が未知の場合、データが対になっており、対応するデータとなっているとき、分散の推定値のバラツキをより小さく合理的に推定する方式が考えられるか。

(6)　分散が未知の場合、データが対になっていなければ、2 つの母集団の分散をそれぞれの標本からどのように推定するか。

以上、本小冊子ではこれらの事柄について詳細に述べてゆくことにする。

解説にあたり始めに、平均の差についての統計的検定の基礎について再掲しておこう。

正規母集団 $N(\mu_x, \sigma_x^2)$ からの独立な m 個の標本を $(x_1, x_2, \cdots, x_m)$、正規母集団 $N(\mu_y, \sigma_y^2)$ からの n 個の独立な標本を $(y_1, y_2, \cdots, y_n)$ と書く。また x_i, y_j は独立であるとする。

$$
\begin{aligned}
\bar{x} &= \sum_{i=1}^{m} x_i/m \qquad x \text{ の標本平均} \\
\bar{y} &= \sum_{j=1}^{n} y_i/n \qquad y \text{ の標本平均} \\
S_{xx} &= \sum_{i=1}^{m} (x_i - \bar{x})^2 \qquad x \text{ の平方和}
\end{aligned}
$$

$$
\begin{aligned}
S_{yy} &= \sum_{j=1}^{n}(y_j-\bar{y})^2 && y \text{ の平方和} \\
S &= \sum_{i=1}^{m}(x_i-\bar{x})^2+\sum_{j=1}^{n}(y_j-\bar{y})^2 && x \text{ と } y \text{ を合わせた平方和} \\
V_x &= S_{xx}/(m-1) && x \text{ の標本分散（} x \text{ の分散の不偏推定量）} \\
V_y &= S_{yy}/(n-1) && y \text{ の標本分散（} y \text{ の分散の不偏推定量）} \\
V &= S/(m+n-2) && x \text{ と } y \text{ を合わせた標本分散（} x\text{、}y \text{ の分散の不偏推定量）}
\end{aligned}
$$

これら $\bar{x}$、$\bar{y}$ は平均、および V_x、V_y、V は分散の最小不偏分散推定量になっている。

統計量、$\bar{x}-\bar{y}$ の基準化は以下のように計算する。
$E(\bar{x}-\bar{y})=\mu_x-\mu_y$
$V(\bar{x}-\bar{y})=\sigma_x^2/m+\sigma_y^2/n$ であるから、

$$z=\frac{\{(\bar{x}-\bar{y})-(\mu_x-\mu_y)\}}{\sqrt{\frac{\sigma_x^2}{m}+\frac{\sigma_y^2}{n}}}$$

すなわち、統計量からその期待値を引き、標準偏差 (分散の平方根) で割ることにより、基準化された統計量 z が得られ、x、y が正規分布に従っていれば z は標準正規分布に従う。
もしも分散が未知であれば、それらを不偏推定量 $\hat{\sigma}^2$ で置き換えれば、

$$t=\frac{\{(\bar{x}-\bar{y})-(\mu_x-\mu_y)\}}{\sqrt{\frac{\hat{\sigma}_x^2}{m}+\frac{\hat{\sigma}_y^2}{n}}}=\frac{\{(\bar{x}-\bar{y})-(\mu_x-\mu_y)\}}{\sqrt{\frac{V_x}{m}+\frac{V_y}{n}}}$$

は近似的に自由度 ϕ の t 分布に従う ($Welch$ の検定)。
$\sigma_x^2=\sigma_y^2=\sigma^2$ と仮定できるときは、

$$t=\frac{\{(\bar{x}-\bar{y})-(\mu_x-\mu_y)\}}{\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)V}}$$

が自由度 $(m+n-2)$ の t 分布に従う。

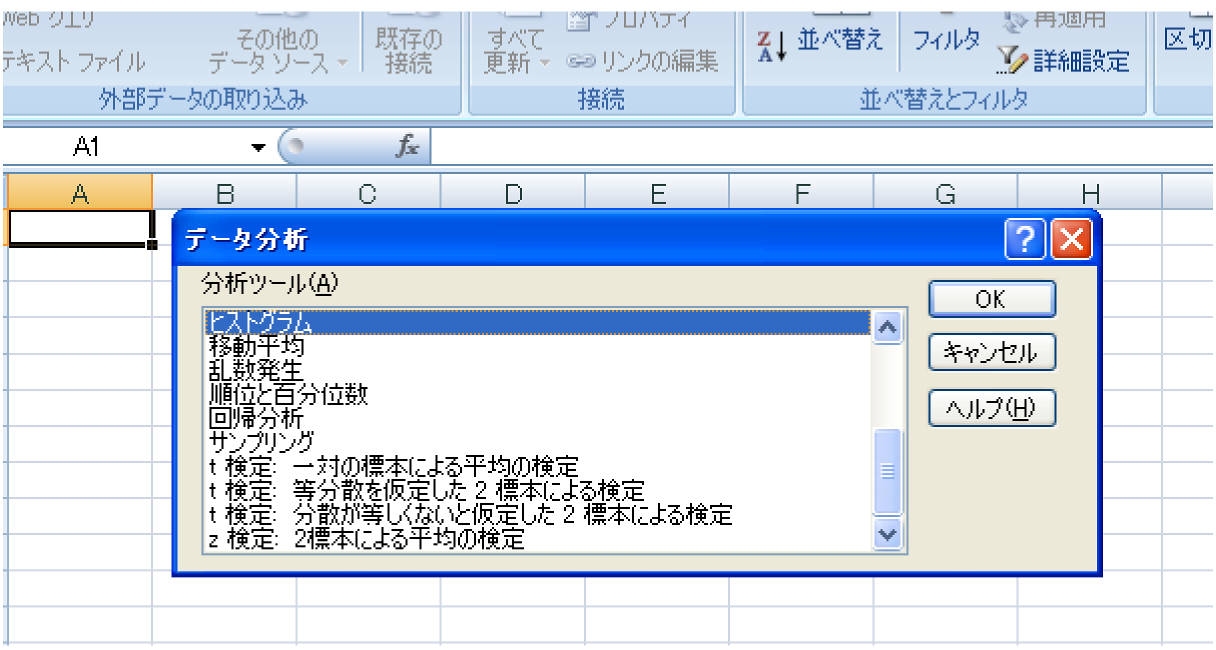

図 1. エクセルにおける、分析ツールのメニュー

平均の差に関係するエクセルにおける「分析ツール」には次のようなものが用意されている。(分析ツールを初めて使用するためには、「ツール」、「アドイン」から「分析ツール」をチェックして「ツール」のメニュー上で「分析ツール」を使用可能にする。エクセル2007以降では「オフィスボタン」から「エクセルのオプション」、「アドイン」、「詳細設定」で「分析ツール」をチェックし元へ戻ると、メニュー「データ」タブの一番右端に「分析ツール」というメニューが現れる)

(1)　t 検定：一対の標本による平均の検定

使用前、使用後というような対となるデータの平均の差を検定する手法である。この手法は個体間の違いを相殺するので、単なる2標本による平均の検定より検出力 (仮説が真で無いとき、仮説が棄てられる確率) が大きいことが期待される。

(2)　t 検定：等分散を仮定した2標本による検定

この手法は分散が未知の場合で、従来最もよく使われて来た方法で、平均の差の検定と言えば t 検定としてこれが使われて来た。しかし、最近では x または y の分布が非対称であったり、明らかに分散の大きさが違っている場合があるのでこの検定法を使用するのに注意が必要となってきた。

(3)　t 検定：分散が等しくない2標本の検定

最近の実験では標本数もそれほど多く取れず、等分散の仮定もはっきり示されない場合が多く、分散未知のときは安全を考えてこの検定法 ($Welch$ の検定ともいう) を使うケースが多くなってきている。ただし、近似自由度の計算式が非常に複雑である。

(4)　z 検定：2標本による平均の検定

正規分布に基づく分散が既知の場合の2標本による平均の差の検定法である。標準正規分布を表わす記号として z を使っている。

(5)　F 検定：2標本を使った分散の検定

$Welch$ の検定などを使用する場合など、等分散であることの証明は出来ないが、「等分散でないことを否定できない」ことを主張するための検定である。2標本のうち、分散の比を計算するにあたり、分母に小さい方、分子に大きい方をとり計算するので、従って必ず両側 F 検定になる。

２．分散が既知のときの平均の差の検定

帰無仮説 $H_0 : \mu_x = \mu_y$ の下では、

$E(\bar{x} - \bar{y}) = 0$

$V(\bar{x} - \bar{y}) = \sigma_x^2 + \sigma_y^2$ である。これを基準化すると、

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}}}$$

であるから両側検定では、$|z|$ が正規分布の 100α パーセント点、$K_{\alpha/2}$ より大のとき仮説が棄却される。

片側検定では、上側対立仮説に対しては $z > K_\alpha$、もしくは下側対立仮説に対しては $z < -K_\alpha$ のとき、仮説が棄却される。

いずれにせよ z の絶対値が限界点を超えたとき仮説が棄却される、と覚えておけばよい。なお、$\sigma_x^2 = \sigma_y^2 = \sigma^2$ のとき、z の式はより簡単に書けて、

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)\sigma^2}}$$

で計算する。
ここでいう検定方式はエクセルでは (4)z 検定：2 標本による平均の検定 z 検定がそれで、

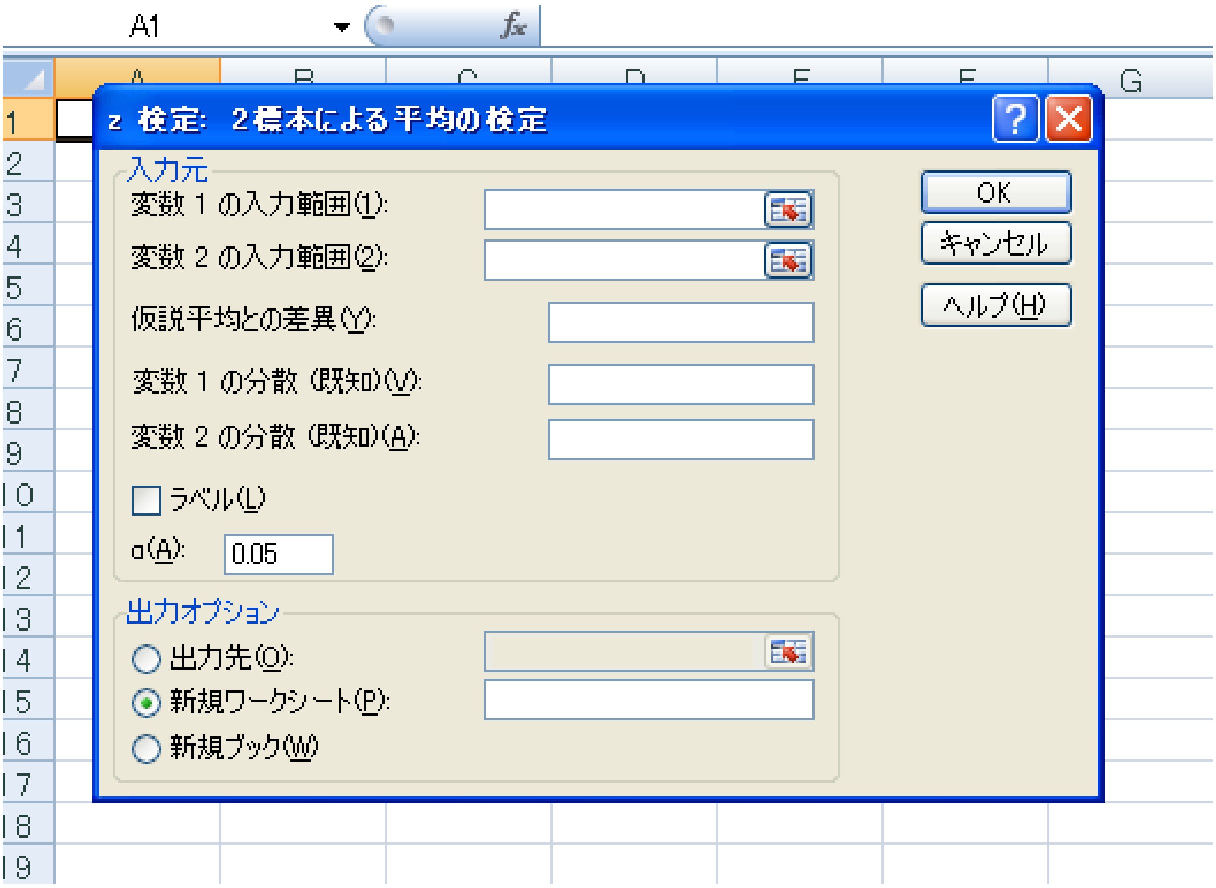

図 2. エクセルの分析ツール、z-検定：2 標本による平均の検定

標本 1 と標本 2 のセル範囲を、変数 1 の入力範囲 (1)、変数 2 の入力範囲 (2) に入れる。このとき先頭行にラベルがある場合には、ラベルのある先頭行から範囲を指定し、かつ□ラベルの枠にチェックを入れる。仮説平均との差異には $H_0 : \mu_x = \mu_y$ のときは 0 を入れる。変数 1 の分散、変数 2 の分散には既知の値を入れ、出力オプションは◯出力先、または◯新規ワークシートのいずれを選択しても良い。ただし、出力先を選んだときは、直ちに右枠をクリックして出力先のセル番地を入れなければならない。このオプションは現在使用中のシートに出力するためのものである。なお、検定の水準はデフォルトで 0.05 になっているが、1%検定にしたければ 0.01 を指定すればよい。
【例題】次の表は A 社、B 社製それぞれのナイロン樹脂の強度を測定したものである。このデータを基に両社の製品の強度に違いがあるかどうか、検定の水準 5%で検定を行う。

A社	78.5	78.6	78.2	79.1	78.4	78.9	80.9	79.0	79.1
B社	75.3	75.9	73.2	75.1	73.1	75.0	75.4	73.0	

ただし、長年の経験から製品の分散は A 社が 0.70、B 社が 1.50 であるとする。
この例題の結果は次図で示す通りである。
　出力先はこのシートの $\$A\$12$ を指定したので表のすぐ下に結果が表示される。正規化された値は 8.672276 となり、この値が正規分布の両側パーセント点 1.959964 より大きいので、検定結果は「帰無仮説は棄却される」。すなわち両社の製品に違いが存在するという結論に

なる。片側検定であれば限界値が 1.644854 であり、その p 値はいずれも 0.0000 であることが表から読み取れる。

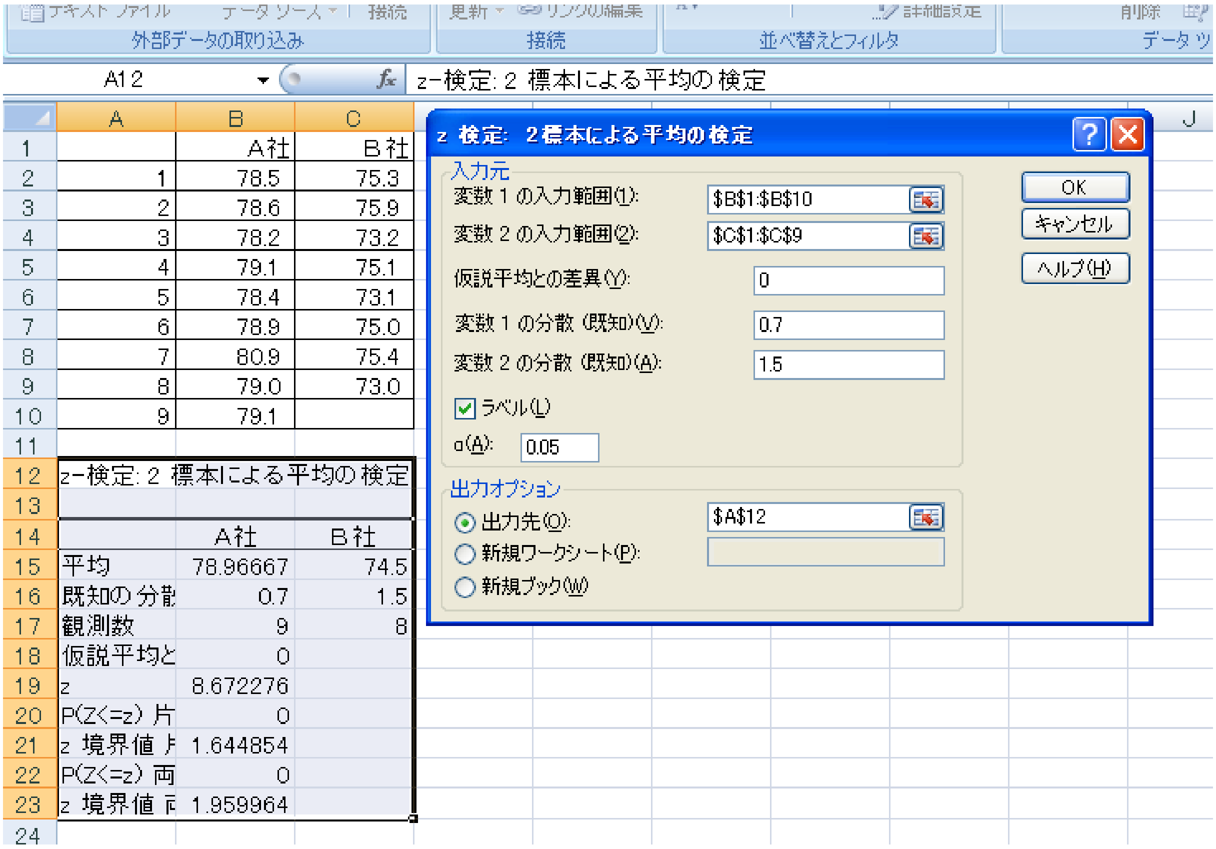

図 3.　z-検定による例題の結果

3．分散が未知のとき (母集団の分散が等しいと仮定) の平均の差の検定

2 つの標本の母集団の分散が既知の場合は実際には滅多になく、ほとんどの場合未知である場合が多く、さらにこの母集団の分散が等しいと仮定してよいという確証も得られないのが普通である。しかし、この理論の簡便さから、これまで長い間この検定法が使われて来たのは紛れもない事実で、2 つの母集団の平均の差の検定といえば分散が等しいと仮定した場合の t 検定のことを意味していた。近年になって分散が異なる場合の近似検定 ($Welch$ の検定) が知られるようになってからは、この t 検定は使われないようになってきた。

ところで、$\sigma_x^2 = \sigma_y^2 = \sigma^2$ のとき、z の式は次式のようになり標準正規分布に従うが、

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)\sigma^2}}$$

と書いたとき、この σ^2 が未知なので、2 つの標本の母集団の分散が等しい時の分散の推定量の最小不偏分散の分散 V を計算すると、

$$S = \sum_{i=1}^{m}(x_i - \bar{x})^2 + \sum_{j=1}^{n}(y_j - \bar{y})^2$$

$$V = S/(m + n - 2)$$

この共通分散 V を使って基準化すると、t は自由度 $(m + n - 2)$ の t 分布に従うから、

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)V}}$$

となり、この値の絶対値が t 分布の限界値 $t(\alpha, m+n-2)$ または $t(\alpha/2, m+n-2)$ を超えるとき、仮説 H_0 は棄てられる。すなわち、検定で有意であると判定される。
【例題】先の例題で、分散未知、2 つの標本の母集団の分散が等しいと仮定してもよいときの、平均の差の検定を行ってみる。

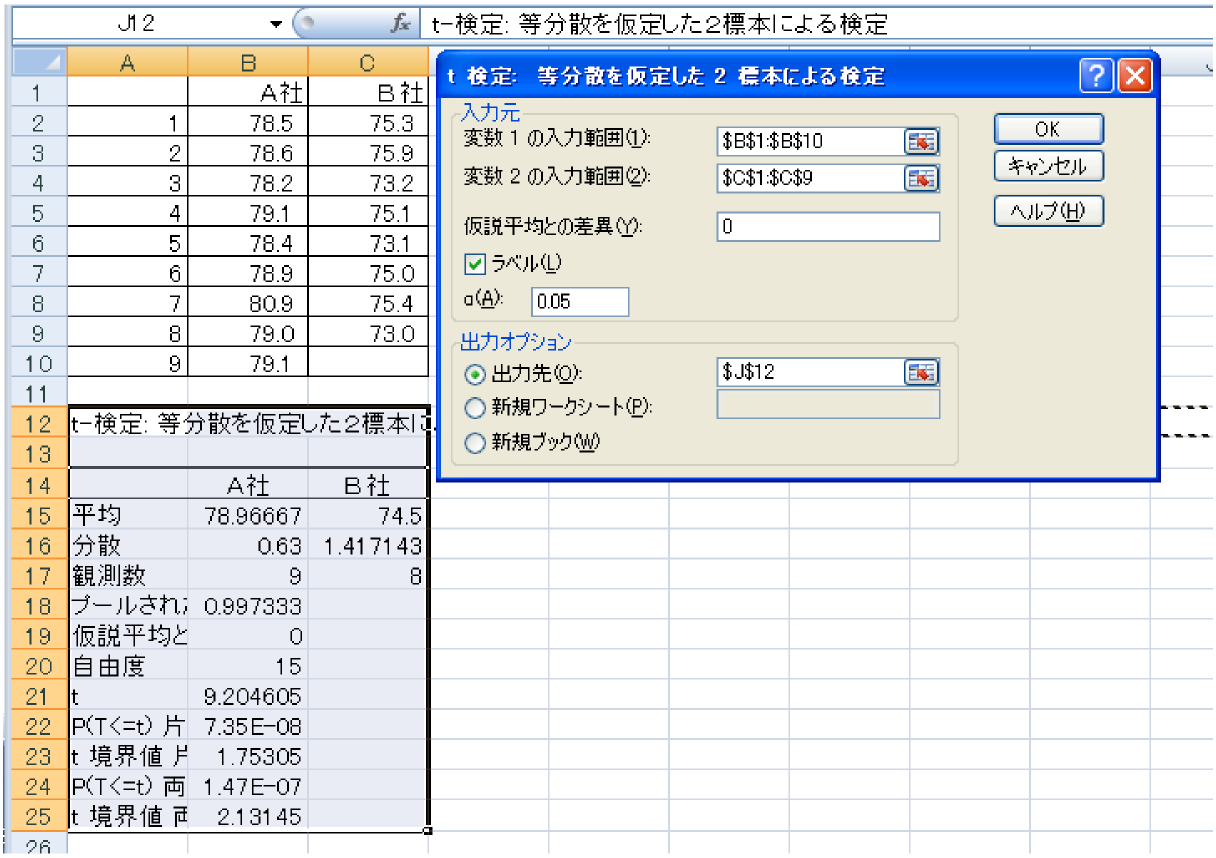

図 4.　エクセルでの t 検定：等分散を仮定した 2 標本による検定

共通分散 V はプールされた分散として、0.997333、仮説平均の差は当然 0、自由度は m+n-2 であるから 15、計算された t 値は 9.204605、片側 5%限界値は 1.75305、両側限界値は 2.13145 であるからこの検定は有意であるという結論になる。このときの p 値は片側検定で $7.35 * 10^{-8}$、両側検定で $1.47 * 10^{-7}$ となり、いずれもほとんど 0 であるといえる。

4．分散が未知のとき：一対の標本による平均の検定

この検定法は 2 標本で「一対の平均の差の検定」ともいわれ、従来の t 検定に比べ、1 対のデータを用いるので個体間のバラツキが相殺され、より違い (差) がはっきりする検定法であるといえる。

測定値＝真値＋誤差

誤差＝個体間誤差＋個体内誤差＋測定誤差

$$\sigma^2 = \sigma_B^2 + \sigma_W^2 + \sigma_M^2$$

一般に、

$V(x-y) = V(x) + V(y) = 2\sigma^2$　　(ただし、$V(x) = V(y) = \sigma^2$、かつ x と y は独立とする)

ところで、個体 i に対してペアで観測値 $(x_i, y_i), i = 1, 2, \cdots, n$ が得られたとして、その差

$x_i - y_i$ についての分散は、

$$V(x_i - y_i) = \sigma_P^2 = 2\sigma_W^2 + 2\sigma_M^2$$ であり、

$$V(x - y) > V(x_i - y_i) \quad \Leftrightarrow \quad 2\sigma^2 > \sigma_P^2$$

であることは分かるが、個体間分散が未知であるために σ_P^2 の大きさを知ることが出来ない。そこで、一対の観測値 $(x_i, y_i), i = 1, 2, \cdots, n$ が得られたとして、

$H_0 : \mu_x - \mu_y = 0$

$H_1 : \mu_x \neq \mu_y$ （$\mu_x > \mu_y$, または$\mu_x < \mu_y$でもよい。この場合は片側検定になる。）

検定統計量は $\bar{x} - \bar{y}$ であり、

$E(\bar{x} - \bar{y}) = 0$

$V(\bar{x} - \bar{y}) = \dfrac{\sigma_P^2}{n}$

従って、H_0 の下では、

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{V_P}{n}}} \quad \sim \quad t_{n-1}$$ 自由度 $n-1$ の t 分布に従う。

ただし、

$$\hat{\sigma}_P^2 = V_P = \frac{1}{n-1}\sum_{i=1}^{n}\{(x_i - y_i) - (\bar{x} - \bar{y})\}^2$$

で計算される。

このとき、次のことに注意しよう。

$$\sum_{i=1}^{n}\{(x_i - y_i) - (\bar{x} - \bar{y})\}^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{n}(y_i - \bar{y})^2 - 2\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$
$$= S_{xx} + S_{yy} - 2S_{xy} = (n-1)\{V(x) + V(y) - 2Cov(x, y)\}$$

同一個体の差をとっているので、S_{xy} や $Cov(x, y)$ は正であるから、対応のある場合の平均の差の分散は必ず対応がない場合の分散以下になることが分かる。

【例】以下のデータはある中学校女子の立ち幅跳びのデータで、「前」、後」はある訓練の後に計測して、訓練前の跳んだ距離との結果を比較した表である。このデータからある訓練が立ち幅跳びの距離に影響を与えたかどうか有意水準 5% で検定してみる。

立ち幅跳びの訓練前と訓練後のデータ ($n = 10$)

	1	2	3	4	5	6	7	8	9	10
前 (y_i)	195	145	145	128	163	165	154	159	156	141
後 (x_i)	202	167	161	145	164	166	180	168	166	170

訓練前と後の差についてその個々の差の平均を計算してみる。この場合前後の平均の差に一致し、

$$average(x_i - y_i) = \frac{1}{10}\sum_{i=1}^{10}(x_i - y_i) = \bar{x} - \bar{y} = 13.8$$

そして $(x_i - y_i), i = 1, 2, \cdots, 10$ の分散は、

$$V_P = \frac{1}{9}\sum_{i=1}^{10}\{(x_i - y_i) - 13.8\}^2 = 97.06666667$$

であるから、

$$t_0 = \frac{\bar{x} - \bar{y}}{\sqrt{V_P/n}} = \frac{13.8}{\sqrt{97.06666667/10}} = 4.429391097$$

これが自由度 $(n-1)=9$ の t 分布に従うことから、t_0 が t 分布の両側 5%点 2.262157158 より大きいので、この検定で有意である、つまり訓練効果があるといえる。

　このことをエクセルの分析ツールから「t 検定:一対の標本による平均の検定」を実行した結果を掲げてみよう。

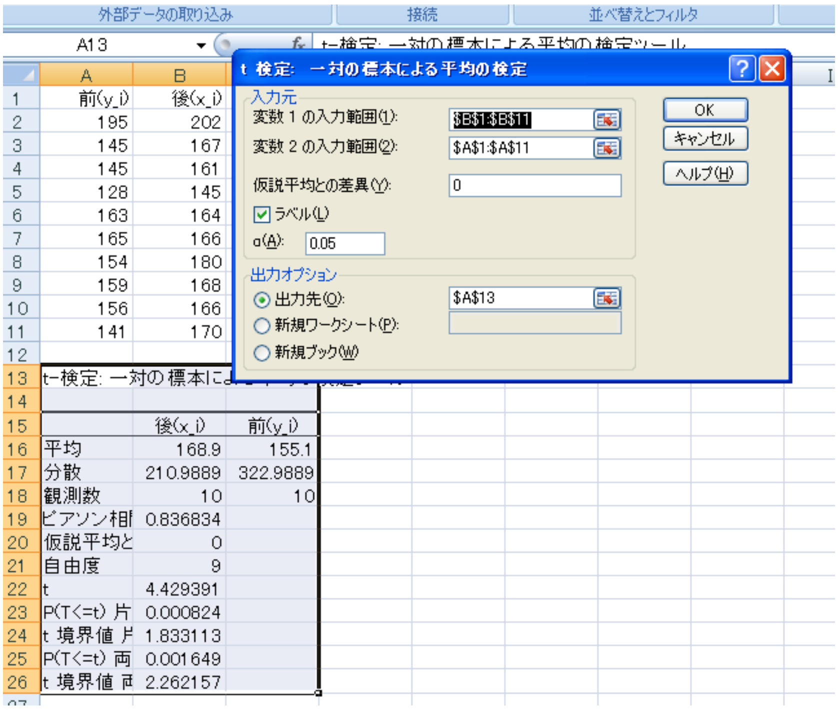

図 5. エクセルでの t 検定：一対の標本による検定

ちなみに、リハビリの前後で分散が等しいとして、一対の標本として扱わない場合の t 検定による平均の差の検定をしてみりと、

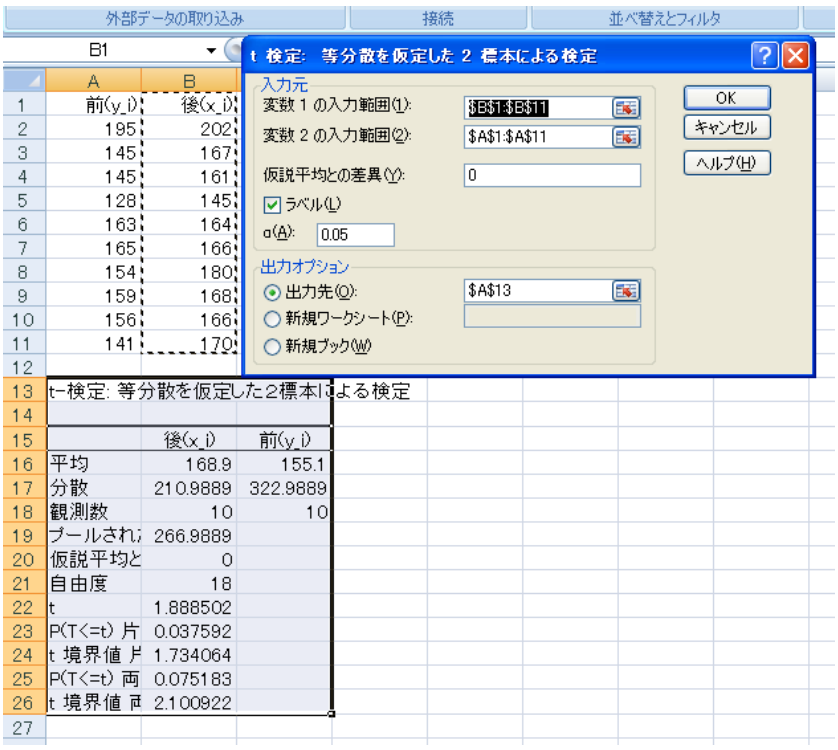

図 6. エクセルでの t 検定：等分散を仮定した場合の検定

５．分散が未知のとき：分散が異なるときの平均の検定

($Welch$ の検定)

　まず、等分散の仮定が成立しない状況であるかを調べてみる必要がある。
検定の性質から、等分散であるという証明はできないが、少なくとも等分散であるという仮説が否定されねばならない (仮説が棄却されること)。

●等分散の検定

$$x_1, x_2, \cdots, x_m \quad \sim \ N(\mu_x, \sigma_x^2) \quad \Rightarrow \quad \bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_i, \quad V_x = \frac{1}{m-1}\sum_{i=1}^{m}(x_i - \bar{x})^2$$

$$y_1, y_2, \cdots, y_n \quad \sim \ N(\mu_x, \sigma_x^2) \quad \Rightarrow \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \quad V_y = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

このとき、帰無仮説は x、y 両母集団の分散は等しいという仮定であり、つまり $H_0 : \sigma_x^2 = \sigma_y^2$、$H_1 : \sigma_x^2 \neq \sigma_y^2$ という仮説検定となり、従って両側検定になる。

検定統計量は、$F_{n-1}^{m-1} = \dfrac{V_x}{V_y}$ もしくは、$F_{m-1}^{n-1} = \dfrac{V_y}{V_x}$ のどちらか大きい方を F とする。

　$F = max(F_{n-1}^{m-1}, F_{m-1}^{n-1})$,　　ここで、$max(a, b)$ は a と b の大きい方をとる記号法

このことから F は必ず 1 より大になることが分かる。

　F は自由度 (ϕ_1, ϕ_2) の F 分布 (ϕ_1 は分子、ϕ_2 は分母の自由度) に従うので、

　$F \geq F_{\phi 2}^{\phi 1}(0.025)$,　$F \geq F_{\phi 2}^{\phi 1}(0.005)$ のとき、

F 検定は有意となる。すなわち仮説は棄却され、2 つの母集団の分散は等しいという仮説は成り立たないということになる。それゆえ以下に述べるような特別な検定方式が必要となる。

●分散が未知で、異なる場合の 2 つの母集団の平均の差の検定

　　（これは通称、$Welch$ の検定と呼ばれている）

$H_0 : \mu_x - \mu_y = 0$

$H_1 : \mu_x \neq \mu_y$　($\mu_x > \mu_y$, または $\mu_x < \mu_y$ でもよい。この場合は片側検定になる。)

検定統計量は $\bar{x} - \bar{y}$ であり、

$E(\bar{x} - \bar{y}) = 0$

$$V(\bar{x} - \bar{y}) = \frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}$$

従って、H_0 の下では、

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{V_x}{m} + \frac{V_y}{n}}} \qquad \sim \qquad t_\phi$$

近似的に自由度 ϕ の t 分布に従うものと仮定する。

ただし、この自由度 ϕ の計算が厄介であり、次式で示す複雑な式 ($Satterthwaite$ による) になる。

$$\phi = \frac{\left(\frac{V_x}{m} + \frac{V_y}{n}\right)^2}{\left(\frac{V_x}{m}\right)^2/(m-1) + \left(\frac{V_y}{n}\right)^2/(n-1)}$$

で計算される。従って多くの場合 ϕ は小数点のついた自由度になる。勿論補間式を使ってより精度の高い t の限界値を求めても良いが、四捨五入による整数自由度の求め方や、小数点以下を切り捨てて求める方法などがある。しかし安全側を考えれば少し大きな t の限界値を採用する後者の方が良いと思われる。

【例】前ページと同様の問題で考えてみよう。

このデータからある訓練が立ち幅跳びの距離に影響を与えたかどうか有意水準 5% で検定し

てみる。

立ち幅跳びの訓練前と訓練後のデータ $(n = 10)$

	1	2	3	4	5	6	7	8	9	10
前 (y)	195	145	145	128	163	165	154	159	156	141
後 (x)	202	167	161	145	164	166	180	168	166	170

まず基本統計量は、$\bar{x} = 168.9$、$\bar{y} = 155.1$　　$V_x = 210.9888889$、$V_y = 322.9888889$
ここで、$\sigma_x^2 = \sigma_y^2 = \sigma^2$ として、共通の分散を計算すると、$\hat{\sigma}^2 = 266.9888889$ であるから、分散が未知で等しい場合の検定は

$$t_0 = \frac{(\bar{x} - \bar{y})}{\sqrt{(1/m + 1/n)\hat{\sigma}^2}} = \frac{(168.9 - 155.1)}{\sqrt{(1/10 + 1/10) \cdot 266.9888889}} = 1.888502198$$

自由度は $\phi = m + n - 2 = 10 + 10 - 2 = 18$ だから $t_{18}(\frac{0.05}{2}) = 2.100922037$ に比べて t_0 は小さいので、この検定結果は有意ではない。仮説は棄てられない、すなわち訓練効果があるとはいいきれない。

さて $\sigma_x^2 = \sigma_y^2 = \sigma^2$ なる仮定は正しい判断であるかどうか分散の検定で確認してみよう。

$$F_0 = \frac{V_y}{V_x} = \frac{322.9888889}{210.9888889} = 1.530833641$$

は $F(9, 9, 0.05/2) = 4.025994158$ より小さいので 5% 検定で有意であるとはいえない。

さて、本題に入ろう。$\sigma_x^2 \neq \sigma_y^2$ であるから t 統計量は Welch の検定法により、

$$t_0 = \frac{(\bar{x} - \bar{y})}{\sqrt{(V_x/m + V_y/n)}} = \frac{(168.9 - 155.1)}{\sqrt{(210.9888889/10 + 322.9888889/10)}} = 1.888502198$$

この例では $m = n = 10$ であるので $\sigma_x^2 = \sigma_y^2 = \sigma^2$ の場合と同じ計算結果になる。ただし、自由度が $m + n - 2$ とならず、計算式により、ϕ を求めることになる。その結果、$V_x/m = 21.09888889$、$V_y/n = 32.29888889$
を使って、

$$\phi = \frac{\left(\frac{V_x}{m} + \frac{V_y}{n}\right)^2}{\left(\frac{V_x}{m}\right)^2/(m-1) + \left(\frac{V_y}{n}\right)^2/(n-1)} = \frac{(21.09888889 + 32.29888889)^2}{21.09888889^2/9 + 32.29888889^2/9} = 17.24148471$$

であるからそれを切り捨てて、自由度は 17 とする。

自由度 17 のときの t 分布の両側 5% 点は、$t_{17}(\frac{0.05}{2}) = 2.109815559$ であるから、Welch の検定統計量 $t_0 = 1.888502198$ は勿論これより小さいので有意とはならない。

ϕ の式において、$V_x = V_y = V$ とおいて $m + n =$ 一定として最大値を求めると、$m = n$ のとき最大になり、$\phi = 2(m-1) = 2(n-1)$ となり、2 つの母分散を等しいと仮定したときの t 検定に一致する。また $n = m$ としたとき、$V_x = \alpha V$、$V_y = (1-\alpha)V$ で計算すると $\alpha = 0.5$ のとき最大値 $\phi = 2(m-1) = 2(n-1)$ となり、やはり等分散のときが最大の自由度となる。Welch の検定では 2 つの母集団からの分散や自由度が極端にアンバランスなとき、近似自由度が等分散の仮定の自由度からどのくらい小さくなるか注目することになる。すなわち同じ条件下で t 検定をした場合、自由度が大きければ棄却域の限界点はより小さくなるので、帰無仮説が棄却されやすくなり、アンバランスであれば自由度が小さくなるので、逆に限界点が大きくなり帰無仮説が棄却されにくくなる。

以上で得た知見を纏めると、$\sigma_x^2 = \sigma_y^2 = \sigma^2$ かどうかは神様しか知らないので正しい判断は出来かねるが、人間が制御出来るサンプル数はどんな場合にも出来るだけ等しく取った方が効率が良いことが分かる。