Ⅳ-2-2. 和の分散、差の分散

対になったデータの t 検定 (paired t test) というのは、右手と左手とどちらが長いかとか、手と足とどちらが長いかとか、一つの鉢に違う植物をうえて、これを繰り返してどちらの植物の成長が早いかなど、比べる相手が１対１に決まっている場合に行う検定です。この場合、一対のデータの差という２次的データを作るのは簡単です。１対のデータ間の差を１つデータとして、その分散を差の分散として使うことができます。しかし、実際には、Ａという肥料で栽培した作物の収穫量と、Ｂという肥料で栽培した時の収穫量を比較するというように、必ずしも対になっていないことが多いでしょう。同じ鉢に違う肥料を入れたら混ざってしまって実験になりません。そんな場合には、２群のデータを合成して１群のデータセットを作り、その分散を利用して検定を行います。また、反対に 1 つのデータに含まれている複数の要因を抽出し、それらの要因の影響を単独に論じなければならないこともあります。そこで、複数群のデータを足し合わせたり、引いたり、分割することを考えます。また、データとデータを掛け合わせたものを確率変数としてその分布を考えることもあります。これは、主として回帰分析に使われるので、ここで扱おうとするグループ間の有意差検定の範囲を超えていますが、データの取り扱いという意味ではまとめて説明したほうが簡単なので、ここで説明します。

Ⅳ-2-2-1. 和の分散

検討するモデル

Ａというデータ群（A_1、A_2，・・・A_m）とＢというデータ群（B_1， B_2・・B_n）があることにします。そこから各データを足し合わせたＡ＋Ｂというデータ群を作るとか、Ａ－ＢやＡ×Ｂを作ることを考えます。対になっている場合にはＡ＋Ｂは（A_1+B_1， A_2+B_2・・・・・・・・・・・A_n+B_n）のように、Ｎ個のデータをつくれば良いのですが、対になっていない場合は、ＡのどのデータとＢのどのデータを足し合わせればよいか決まっていません。そこで、考えられる組み合わせのすべてについて mn 個のデータをつくり、そのデータの分散について考えます。（この解説は原理的な理解のためにやっているので、実用的な意味を考えていません。筆者はデータの和という概念を具体的に使う場面を思いつきません。たとえば、鉢に違う種類の植物を植えて、その成長量の和を確率的に論ずることに何か意味があるとは思いません。ただ、データの和の分散がどのようになるのかを考えることは、分散をいくつかの要因に取り分けるときるときの考え方を理解するのに役立ちます。一度やってみると分散分析の手順を感覚的に理解するのに役立ちます。）

データの和

　わかりやすくするために、m×n の総当り表を作ります（表６）。表 7 はその具体例で、Ａデータ群として(1,5,6)、Ｂのデータ群として(1,5,6,8)を使って、それらの和のデータを作っています。

表６．Ａ，Ｂ総当たりのデータの和

	Ａ１		Ａi		Am	合計	平均
Ｂ１	A_1+B_1		A_i+B_1		A_m+B_1	$\sum_{i=1}^{m}A_i+mB_1$	M_A+B_1
Ｂj	A_1+B_j		A_i+B_j		A_m+B_j	$\sum_{i=1}^{m}A_i+mB_j$	M_A+B_j
Ｂn	A_1+B_n		A_i+B_n		A_m+B_n	$\sum_{i=1}^{m}A_i+mB_n$	M_A+B_n
合計	$nA_1+\sum_{j=1}^{n}B_j$		$nA_i+\sum_{j=1}^{n}B_j$		$nA_m+\sum_{j=1}^{n}B_j$	$n\sum_{i=1}^{m}A_i+m\sum_{j=1}^{n}B_j$	
平均	A_1+M_B		A_i+M_B		A_m+M_B		M_A+M_B

表 7. 具体例を表１の形式で書いたもの

総和と平均

	1	5	6	Sum	Mean
1	2	6	7	15	5
5	6	10	11	27	9
6	7	11	12	30	10
8	9	13	14	36	12
Sum	24	40	44	108	9
Mean	6	10	11	9	

合計が 108 その平均 M=108/12=9

A の平均 M_A＝4　SS_A　＝14　分散$\sigma^2{}_A=7$

Bの平均 M_B＝5 $SS_{B.}$＝26　分散$\sigma^2{}_b=8.66667$

全体の平均を見てください。$M=M_A+M_B$ になっています。データを足し合わせているのだから、その平均値も平均値の和になるというのは当然ですね。元の２つのデータ群の分散を知っているのだから、これらを利用して、足し合わせたデータの母集団の平均値周りの分散を推定するための SS を計算する方法を考えます。この SS をSS_{A+B}、分散をσ_{A+B}と表すことにします。

個々のデータ x をx_i、M を平均値、x_i と表すと e_i は個々のデータの平均値からの隔たりとすると。

$$x_i=M+e_i$$

式 38

標本集団の２次の積率は$\mathrm{E}\{(x-M)^2\}$ですから
データは平均的に$\sqrt{\mathrm{E}\{(x-M)^2\}}$、平均値から隔たっていることになります。

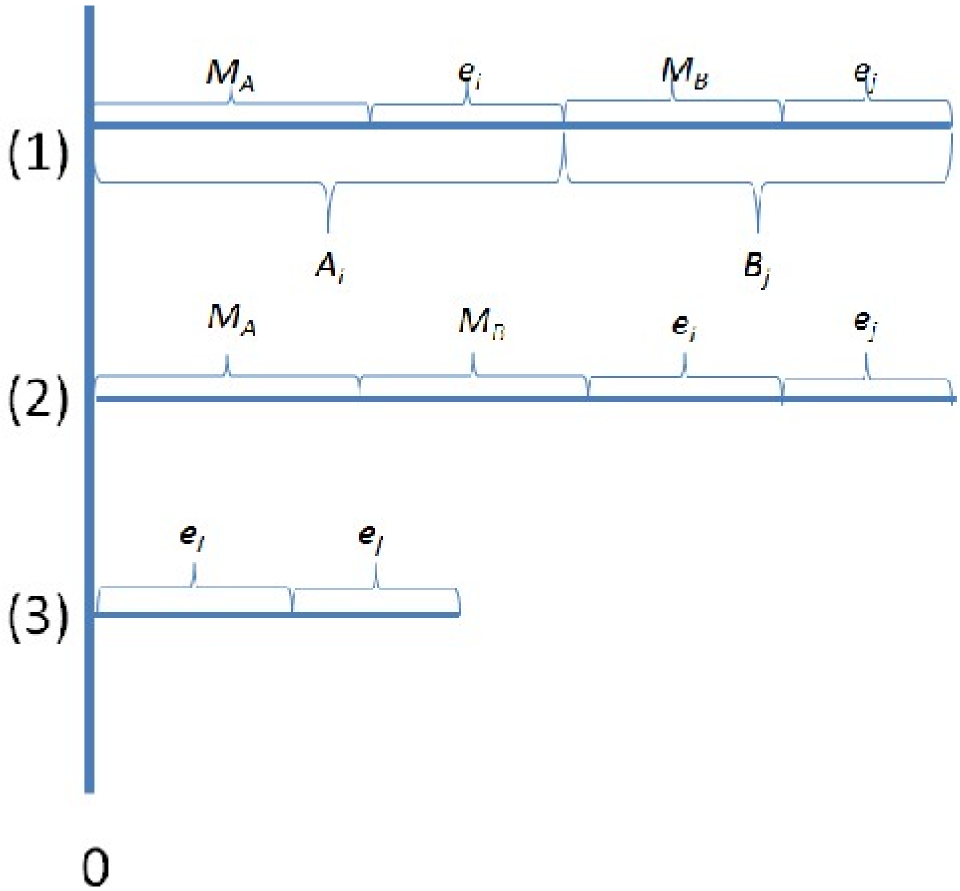

図 32. 合成されたデータの構成

ここで、個々のデータ、A_i+B_j がどのような要素から構成されているかを考えると、図 32 の（１）に示した要素から構成されていることがわかります。図の横棒は A_i+B_jの値を表す数直線です。直線の左端が０です。要素の足し算でできている値ですから、（２）のように順番を入れ替えても値は変わりません。また、M_A 、M_Bは平均値ですから、すべて共通で、ここの値の平均値からの距離（偏差）を考える場合には、M_A 、M_B の値を取り除いて、（３）のように偏差だけを考えればよいことになります。ここでは e_i も e_j も２正の値の例を示しましたが、もちろん、どちらも負の値を取ることがあります。したがって、（３）の例では、０点よりも直線が左に伸びることもあります。
ここで、この足し合わせたものの平均値からの隔たりを$e_{A+B_{ij}}$とあらわすと（持ってまわった言い方ですが、簡単に言えば、（３）の直線の長さのこと)、

$$e_{A+B_{ij}} = e_{A_i} + e_{B_j}$$

表 8. 合成されたデータの偏差の表

	A 1		Ai		Am	合計	平均
B 1	$e_{A_1}+e_{B_1}$		$e_{A_i}+e_{B_1}$		$e_{A_m}+e_{B_1}$	$\sum_{i=1}^{m} e_{A_i}+me_{B_1}$	e_{B_1}
							e_{B_2}
B j	$e_{A_1}+e_{B_j}$		$e_{A_i}+e_{B_j}$		$e_{A_m}+e_{B_j}$	$\sum_{i=1}^{m} e_{A_i}+me_{B_j}$	e_{B_j}
B n	$e_{A_1}+e_{B_n}$	$e_{A_2}+e_{B_n}$	$e_{A_i}+e_{B_n}$		$e_{A_m}+e_{B_n}$	$\sum_{i=1}^{m} e_{A_i}+me_{B_n}$	e_{B_n}
合計	$ne_{A_1}+\sum_{j=1}^{n} e_{B_j}$	$ne_{A_2}+\sum_{j=1}^{n} e_{B_j}$	$ne_{A_i}+\sum_{j=1}^{n} e_{B_j}$		$ne_{A_m}+\sum_{j=1}^{n} e_{B_j}$	0	0
平均	e_{A_1}	e_{A_2}	e_{A_i}		e_{A_m}	0	0

と表せます。これを使って、表 6 を偏差だけの計算式に書き変えたのが表 8 です。
偏差の和ですから、和の欄のシグマ記号のところが 0 になります。合計と平均が 0 になっていることを確認してください。
ここで、足し合わせてできた標本集団の２次の積率（標本分散）を考えます。
2 次の積率とは、平均値から個々のデータの距離の２乗の平均ですね。母集団の分散の推定値（母集団の２次の積率）は、SS/自由度でしたね。標本集団のデータの 2 次の積率
Ｅ（(ｘ－μ）^{2}）は SS/標本数でした。
２乗の総和をデータ数で割って、次の式になります。

$$\frac{\sum_{i=1}^{m}\sum_{j=1}^{n} e_{ij}{}^2}{mn}=\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\left(e_{A_i}+e_{B_i}\right)^2}{mn}$$

この値のルートを開いた値が、個々のデータの平均値からの距離の平均ですね。面倒なので、記述を簡略化します。このようにして得られる平均値を

$$\overline{e_A{}^2}=\frac{\sum_{i=1}^{m}\sum_{j=1}^{n} e_{A_{ij}}{}^2}{mn}$$

のように表します。つまり

$$\overline{e_{ij}{}^2}=\frac{\sum_{i=1}^{m}\sum_{j=1}^{n} e_{ij}{}^2}{mn}$$

そうしてみると
$\sum_{i=1}^{m}\sum_{j=1}^{n} e_{A_i}{}^2$には B の要素が含まれていないので、$\sum_{j=1}^{n} e_{A_i}{}^2$が意味することは同じものを n 回たすことです。ですから、

$$\overline{e_A{}^2} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} e_{A_i}{}^2}{mn} = \frac{n\sum_{i=1}^{m} e_{A_i}{}^2}{mn} = \frac{\sum_{i=1}^{m} e_{A_i}{}^2}{m}$$

同様に

$$\overline{e_B{}^2} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} e_{B_j}{}^2}{mn} = \frac{m\sum_{j=1}^{n} e_{B_j}{}^2}{mn} = \frac{\sum_{j=1}^{n} e_{B_j}{}^2}{m}$$

さて、(式 47)の関係は、それぞれの平均値についても成り立つので

$$\overline{e_{A+B}} = \overline{e_A} + \overline{e_B}$$

となります。

２次の積率では

$$\overline{e_{A+B}{}^2} = \overline{(e_A + e_B)^2} = \overline{e_A{}^2} + 2\overline{e_A e_B} + \overline{e_B{}^2}$$

$\overline{e_A}$、$\overline{e_B}$は偏差の合計で 0 だから、

$$\overline{e_{A+B}{}^2} = \overline{e_A{}^2} + \overline{e_B{}^2}$$

$$\overline{e_{A+B}{}^2} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left(e_{A+B_{ij}}\right)^2 = \frac{SS_{total}}{mn}$$

$\overline{e_{A+B}{}^2}$は偏差の平方値の合計だから、SS_{total}です。SS_{A+B}ではありません。

$$\overline{e_{A+B}}^2 = \frac{SS_{total}}{mn}, \quad \overline{e_A}^2 = \frac{SS_A}{m}, \quad \overline{e_B}^2 = \frac{SS_B}{n}$$

ですから、

$$\frac{SS_{total}}{mn} = \frac{SS_{\hat{A}}}{m} + \frac{SS_{\hat{B}}}{n}$$

$$SS_{total} = nSS_{\hat{A}} + mSS_{\hat{B}}$$

式 39

となります。

具体的な計算は以下の通りです。

もし、表 7 に示した (2, 6, 7, 6, 10,11, 7,11,12,9, 13,14)という、データ群が、分散が明らかな２つのデータ群を足し合わせたものであることを知らなければ、全体の平均から、それぞれのデータを差し引いて、その２乗の総和を求めるという形で、左辺を計算するでしょう。具体的な計算の手順は、表 9 のようになります。

表 9. SS_{total}の計算 2

	A 1		Ai		Am
B 1	$\left(e_{A_1} + e_{B_1}\right)^2$		$\left(e_{A_i} + e_{B_1}\right)^2$		$\left(e_{A_m} + e_{B_1}\right)^2$
B j	$\left(e_{A_1} + e_{B_j}\right)^2$		$\left(e_{A_i} + e_{B_j}\right)^2$		$\left(e_{A_m} + e_{B_j}\right)^2$
B n	$\left(e_{A_1} + e_{B_n}\right)^2$		$\left(e_{A_i} + e_{B_n}\right)^2$		$\left(e_{A_m} + e_{B_n}\right)^2$

表 10. 表 9 の展開

	A 1		Ai		A m	合計
B 1	${e_{A_1}}^2+2e_{A_1}e_{B_1}+{e_{B_1}}^2$		${e_{A_j}}^2+2e_{A_j}e_{B_1}+{e_{B_1}}^2$		${e_{A_m}}^2+2e_{A_m}e_{B_1}+{e_{B_1}}^2$	$SS_{\hat{A}}+m{e_{B_1}}^2$
B j	${e_{A_1}}^2+2e_{A_1}e_{B_j}+{e_{B_j}}^2$		${e_{A_i}}^2+2e_{A_i}e_{B_j}+{e_{B_j}}^2$		${e_{A_m}}^2+2e_{A_m}e_{B_j}+{e_{B_j}}^2$	$SS_{\hat{A}}+m{e_{B_m}}^2$
B n	${e_{A_1}}^2+2e_{A_1}e_{B_n}+{e_{B_n}}^2$		${e_{A_i}}^2+2e_{A_i}e_{B_n}+{e_{B_n}}^2$		${e_{A_m}}^2+2e_{A_m}e_{B_n}+{e_{B_n}}^2$	$SS_{\hat{A}}+m{e_{B_m}}^2$
合計	$n{e_{A_1}}^2+SS_{\hat{B}}$		$n{e_{A_i}}^2+SS_{\hat{B}}$		$n{e_{A_m}}^2+SS_{\hat{B}}$	$nSS_{\hat{A}}+mSS_{\hat{B}}$

黄色の部分の計算は

$$\sum_{i=1}^{m}{e_{A_i}}^2+2e_{B_j}\sum_{i=1}^{m}e_{A_i}+\sum_{i=1}^{m}{e_{B_j}}^2$$

第１項は *SS*、第２項はシグマの部分は偏差の総和で０、第３項は i を含まない数なので

$$SS_{\hat{A}}+m{e_{B_m}}^2$$

となります。

表 7 のデータを具体的に当てはめて SS を計算すると表 11 のようになります。

表 11. SS 計算の具体例

	1	5	6	Sum
1	49*	9	4	62
5	9	1	4	14
6	4	4	9	17
8	0	16	25	41
Sum	62	30	42	134

表 11 に示したように、たとえば、*は（2－９）^{2}＝７^2＝４９と計算していますが、これを $=\{(2-6)+(2-5)\}^2$ と計算しても 49 になります。

黄色の部分については、14+3×16=62

nSS_A+mSS_B は 4×14+3×２６＝134 です。

(2, 6, 7, 6, 10, 11, 7, 11,12, 9, 13,14)という、データから、平均値を求めて、SS を計算してみてください。確かに一致します。

$$SS_{total}=nSS_{\hat{A}}+mSS_{\hat{B}}$$

という式では。要因 A のよる平方和の部分と要因 B による平方和の部分に分けられています。つまり、平方和を２つの部分に分けることができるということです。

式 39 にもどって、ここで、*SS* を何のために計算しているのかを思い出します。母集団の２次の積率の推定値（分散σ^2）を求めているのです。一般には

$$\sigma^2 = \frac{SS}{標本数 - 1}$$

ですね。

私たちが求めている$\sigma_{A+B}{}^2$は、図２４に示した$e_A + e_B$の平方の合計値ではありません。

そこでもう一度、表８を見ます。行B_1の行に注目します。この行の各列の値と平均値の差は、$e_{A_1} + e_{B_1} - e_{B_1}$, $e_{A_2} + e_{B_1} - e_{B_1}$, $\cdots, e_{A_m} + e_{B_1} - e_{B_1}$ ですから、その平方和は

$$\sum_{i=1}^{m} e_{A_i}{}^2 = SS_A$$

です。

これはどの行についても同じですから、行の数はn ですから、この値の総和は、

$$n\sum_{i=1}^{m} e_{A_i}{}^2 = nSS_A$$

ところで、私たちが求めているのは母集団の平均値周りの二次の積率としての分散ですから、$\sigma_{A+B}{}^2$ と記述すべきものです。各行ごとに平均値との差の平方和として、これを求めるとすると、

$$\sigma_{A+B}{}^2 = \frac{SS_A}{m-1} = \sigma_A{}^2$$

という推定の仕方が可能ですが、同じことが列についても言えて、

$$\sigma_{A+B}{}^2 = \frac{SS_{B.}}{n-1} = \sigma_B{}^2$$

表８

	A 1		Ai		Am	合計	平均
B 1	$e_{A_1} + e_{B_1}$		$e_{A_i} + e_{B_1}$		$e_{A_m} + e_{B_1}$	$\sum_{i=1}^{m} e_{A_i} + me_{B_1}$	e_{B_1}
							e_{B_2}
B j	$e_{A_1} + e_{B_j}$		$e_{A_i} + e_{B_j}$		$e_{A_m} + e_{B_j}$	$\sum_{i=1}^{m} e_{A_i} + me_{B_j}$	e_{B_j}
B n	$e_{A_1} + e_{B_n}$	$e_{A_2} + e_{B_n}$	$e_{A_i} + e_{B_n}$		$e_{A_m} + e_{B_n}$	$\sum_{i=1}^{m} e_{A_i} + me_{B_n}$	e_{B_n}
合計	$ne_{A_1} + \sum_{j=1}^{n} e_{B_j}$	$ne_{A_2} + \sum_{j=1}^{n} e_{B_j}$	$ne_{A_i} + \sum_{j=1}^{n} e_{B_j}$		$ne_{A_m} + \sum_{j=1}^{n} e_{B_j}$	0	0
平均	e_{A_1}	e_{A_2}	e_{A_i}		e_{A_m}	0	0

これから、

$$(m-1){\sigma_{A+B}}^2 = SS_A = (m-1){\sigma_A}^2$$
$$(n-1){\sigma_{A+B}}^2 = SS_B = (n-1){\sigma_B}^2$$

上下の式を足すと

$$(m+n-2){\sigma_{A+B}}^2 = SS_A + SS_B = (m-1){\sigma_A}^2 + (n-1){\sigma_B}^2$$
$${\sigma_{A+B}}^2 = \frac{SS_A + SS_B}{m+n-2} = \frac{(m-1){\sigma_A}^2 + (n-1){\sigma_B}^2}{m+n-2}$$

となります。

$${\sigma_{A+B}}^2 = \frac{SS_{A+B}}{df}$$

ですから、

$$SS_{A+B} = SS_A + SS_B$$
$$df = m+n-2$$

となります。結果をまとめると、以下のようになりますが、

$${\sigma_{A+B}}^2 = \frac{(m-1){\sigma_A}^2 + (n-1){\sigma_B}^2}{m+n-2}$$

式 40

この式をよく見ると、この式は自由度で重みをつけた２つの分散の平均になっています。つまり、等分散性を仮定した時点で、

$${\sigma_{A+B}}^2 = {\sigma_A}^2 = {\sigma_B}^2$$

を受け入れているのですが、実際にデータとして、得られる２つの分散は等しくないから、データ数の違いを考慮して、の重み付き平均をとることになるということです。

次に考えたいのは、A+B のサンプルサイズ N です。自由度が$\mathrm{m}+\mathrm{n}-2$だから、サンプルサイズは$\mathrm{m}+\mathrm{n}-1$です。すこし、違和感がありませんか。総データ数はmnです。だとすると、その自由度は$mn-1$のはずです。A+B のサンプルサイズは m+n-1,自由度は m+n-2 です。平均を使うとその都度、自由度が１つ下がるという考えかたを使って、Ａ＋Ｂの自由度は、SS_Aを作るときに１回、SS_Bを作るときに１回、平均化を行っていますから、その都度ごとに、自由度が１つ減って、$(m-1)+(n-1)$で、自由度は $m+n-2$ で、サンプルサイズは $m+n-1$ と考えた。これはこれで良いはずです。全データと言う意味での total と A と B の要因の和と言う意味での A+B を使うことにして、自由度を df で表すとすると

$$df_{total}=mn-1$$
$$df_{A+B}=m+n-2$$

となるので、この自由度の差に相当する分散が理論的にも現実にも存在するはずですが、我々が、想定した和の分散と言うモデルには、それが組み込まれていないのです。その要因は、A の要因と B の要因が重なり合うことによって生ずる要因と言う意味で、交互作用

と言う名前が付けられたり、場合によっては、説明できない要因と言う意味で残差としてあつかわれたりしますが、記号としては$A \times B$のように表します。その自由度は、

$$\mathrm{df}_{A\times} = \mathrm{df}_{total} - df_{A+B} = (mn-1) - (m+n-2) = (m-1)(n-1)$$

です。良く考えてみれば、我々が考えてきたモデルは明らかに不自然です。
表 7 を再掲しました

表 7

総和と平均

	1	5	6	Sum	Mean
1	2	6	7	15	5
5	6	10	11	27	9
6	7	11	12	30	10
8	9	13	14	36	12
Sum	24	40	44	108	9
Mean	6	10	11	9	

要因 A の１の列と５の列の差を見てください。すべての行で４です。５の列と６の列ではすべて１です。行についてもみてみると、すべての列で差が同じです。こんな不自然なデータはありません。つまり、この差を生み出す要因取り除いて０としたから、このモデルでは、その要因がもたらす変動が０なのです。これが、何かを考える必要がありますが、それはＦ検定のところで行います。

IV-2-2-2. 差の分散

上記の議論を応用してデータの差についてその分散を考えます。これは和とは違って実用的な意味があります。ある植物とある植物の成長量に差があるかというのは、意味のある検討です。同じモデルを使います。

表 12. 差のデータのすべての組み合わせ

	A_1		A_i		A_m	合計
B_1	$A_1 - B_1$		$A_i - B_1$		$A_m - B_1$	$\sum_{i=1}^{m} A_i - B_1$
B_j	$A_1 - B_j$		$A_i - B_j$		$A_m - B_j$	$\sum_{i=1}^{m} A_i - B_j$
B_n	$A_1 - B_m$		$A_i - n$		$A_m - B_n$	$\sum_{i=1}^{m} A_i - B_n$
合計	$A_1 - \sum_{j=1}^{n} B_j$		$A_i - \sum_{j=1}^{n} B_j$		$A_m - \sum_{j=1}^{n} B_j$	$n\sum_{i=1}^{m} A_i - \mathrm{m}\sum_{j=1}^{n} B_j$

表 13.　具体的なデータ

総和と平均

	1	5	6	Sum	Mean
1	0	4	5	9	3
5	−4	0	1	−3	−1
6	−5	−1	0	−6	−2
8	−7	−3	−2	−12	−4
Sum	−16	0	4	−12	−1
Mean	−4	0	1	−1	

Ａというデータ群（A_1、A_2，・・・A_m）とＢというデータ群（B_1，B_2・・B_n）があったとします。Ａ群のデータからＢ群のデータを差し引いたＡ－Ｂというデータ群を作ることができます。すべての組み合わせを考えると、表 12 に示すようにサンプルサイズが mn のデータができます。そのデータの分散について考えます。
たとえばＡデータ群として(1,5,6)、Ｂのデータ群として(1,5,6,8)という例について具体的に計算すると以下のとおり(表 13)。

Ａの平均 M_A＝4　SS_A ＝14　分散 σ^2＝7

Ｂの平均 M_B＝5　SS_B ＝26　分散 σ^2＝8,66667

全体の合計が-12、その平均 M=－12/12=－1、$M=M_A-M_B$

そこで全体の分散を考えると

表 14. SS を求める計算

	A_1		A_i		A_m	合計
B_1	$(A_1-B_1-M)^2$		$(A_i-B_1-M)^2$		$(A_m-B_1-M)^2$	$SS_A+me_{B_1}{}^{2*}$
B_j	$(A_1-B_j-M)^2$		$(A_i-B_1-M)^2$		$(A_m-B_1-M)^2$	$SS_A+me_{B_j}{}^2$
B_n	$(A_1-B_n-M)^2$		$(A_i-B_n-M)^2$		$(A_m-B_n-M)^2$	$SS_A+me_{B_n}{}^2$
合計	$ne_{A_1}{}^2+SS_B$		$ne_{A_i}{}^2+SS_B$		$ne_{A_m}{}^2+SS_B$	nSS_A+mSS_B

表 15. 具体的な例での SS の計算

	1	5	6	Sum
1	1**	25	36	62
5	9	1	4	14
6	16	0	1	17
8	36	4	1	41
Sum	62	30	42	134

表 15 の列と行の合計および全体の合計は、表 11 と全く違わないことが確認できます。
Ｂ＝－Ｂとしただけなので、表４-２に示した、各セルの中の第２項の符号が変わるだけで、第２項は合計すると０になってしまうから、この計算が和の計算と同じことになることは直感的にわかります。
ということは、式 40 がなりたちます。

$$\sigma_{A+B}{}^2 = \frac{(m-1)\sigma_A{}^2 + (n-1)\sigma_A{}^2}{m+n-2}$$

$$\sigma_{A-B}{}^2 = \frac{(m-1)\sigma_A{}^2 + (n-1)\sigma_A{}^2}{m+n-2}$$

式 41

同じことですが、丁寧に添え字の符号を変えて書いた方が良いかもしれません。いずれにしても、この式によって、２つの分散を合成した分散はもとまります。
対になっていない t 検定では、この値を分散として使えば良いということです。

ところで、この解説はわざわざ他の教科書論じていないことを論じています。
普通の教科書では、t 検定の説明で、$\sigma_{A\text{-}B}{}^2$ を $\sigma_A{}^2$ と $\sigma_B{}^2$ を自由度で重みをつけた平均として天下り的に

$$\frac{(m-1)\sigma_A{}^2 + (n-1)\sigma_A{}^2}{m+n-2}$$

と与えて、逆向きに式 41 を導き出します。そのほうがはるかに簡単です。著者は、この定義を習った時に、「自由度で重みを付ける。」という言葉が、すぐに理解できなかったし、なぜそうするのかということについて疑問を持ったので、わざわざ遠回りして、式 41 が導き出される背景を代数的に示したのです。
このテキストでは、理解を深めるためにあえて変なことをやっているのです。
もう一つ分かった重大なことがあります。

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\left(e_{A+B_{ij}}\right)^2 = SS_{total}$$

として、求めたSS_{total}は

$$SS_{total} = nSS_A + mSS_B$$

で、各行の平均値と素の行のそれぞれのセルの値の差の平方和として求めたSS_Aの全魚の総和は、nSS_Aだから、

$$mSS_B = SS_{total} - nSS_A$$

として、mSS_Bを求めることが出来ます。これは、計算の簡便化や計算のミスの発見に役に立ちます。

この章で明らかになったことは役に立つので覚えておきましょう。

１．全平方和は部分平方和の和である。

２．全自由度は部分自由度の和である

３．$\sigma_{A-B}{}^2 = \sigma_{A+B}{}^2 = \frac{(m-1)\sigma_A{}^2 + (n-1)\sigma_A{}^2}{m+n-2}$