確率と最小二乗法による土地の境界復元

誤差の基礎Ⅱ（２変数）

誤差の基礎Ⅰでは一変数の標準偏差、正規分布、確率について説明しましたがここからは座標値、二変数の標準偏差、正規分布、確率について考えてみました。

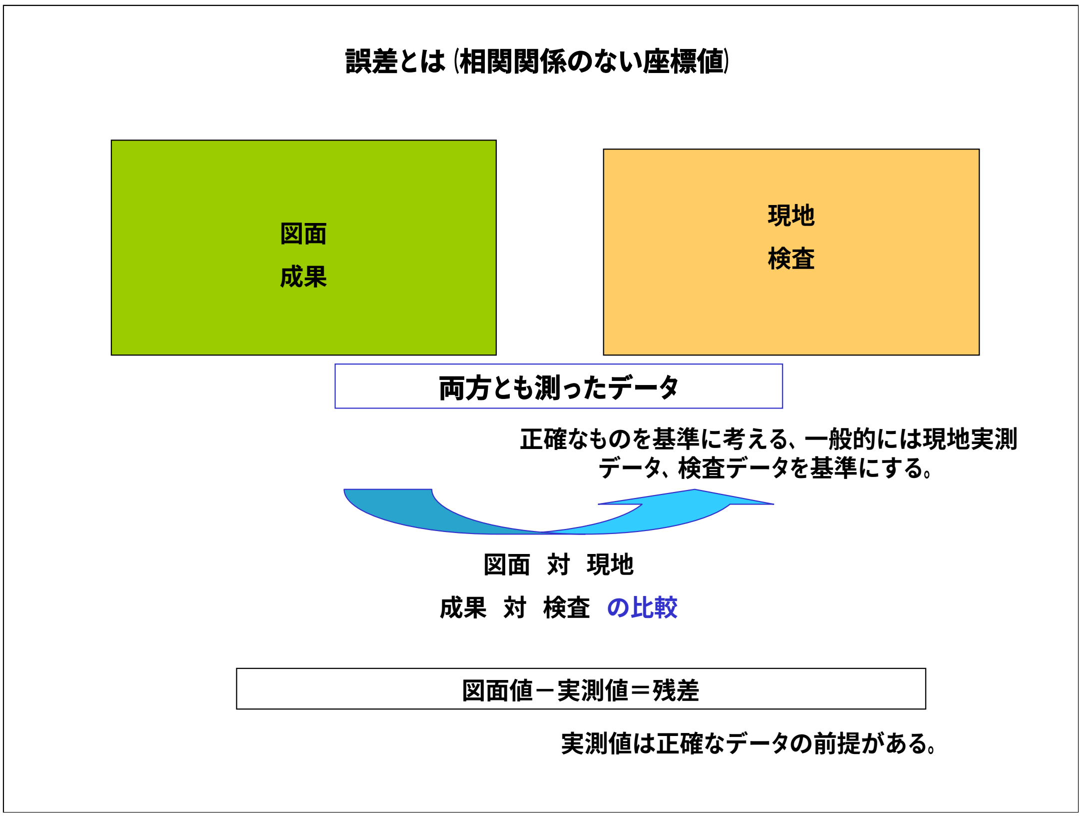
考えて見れば一変数でややっこしいのにと思われるでしょうが結局は二変数をマスターしないと点の位置誤差の議論は出来ません。

二変数の分布はｘとｙの変数で表しますがｘとｙが相違間関係にあるデータと相違間関係にないデータに分けられます。

　ｘを身長、ｙを体重とすれば身長が長ければ体重も重くなる、このような関係を相関関係があると言います。

点の座標値の場合、ｘが増えてもｙはそれに連動して増えません、この関係を相違間関係がないといいます。

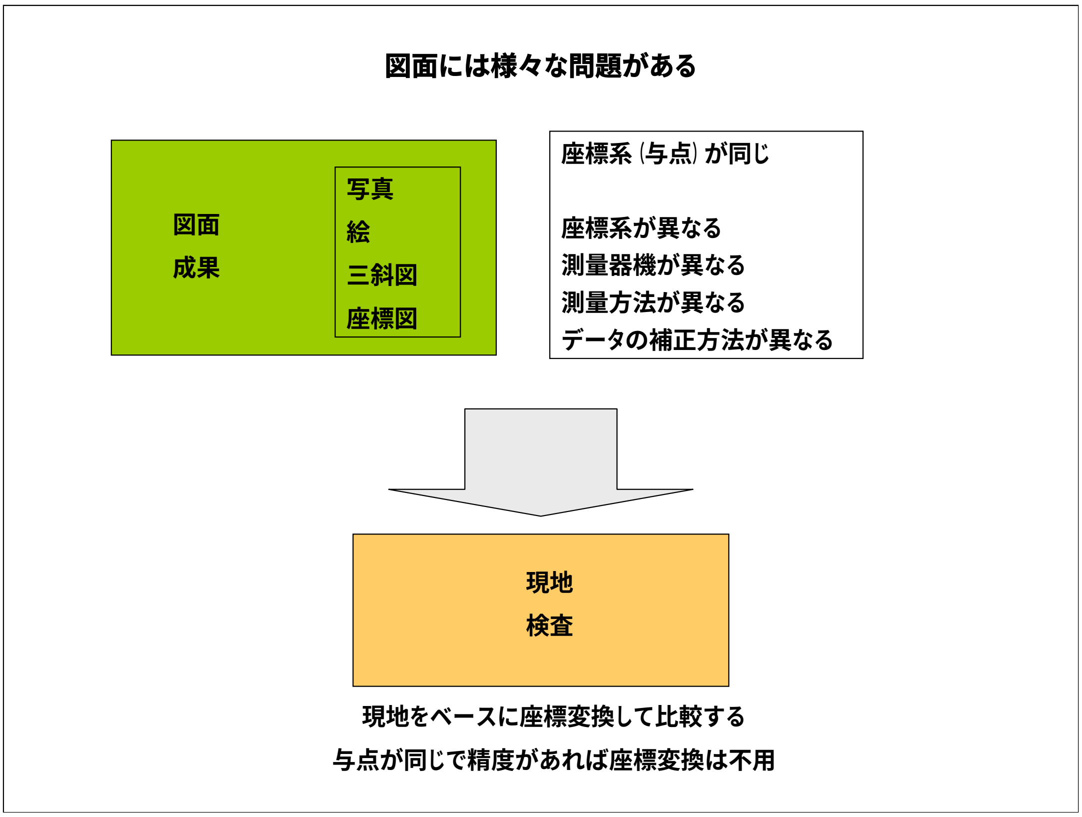
これからの説明は相関関係のないｘとｙの二変数の話です。



誤差とは、一変数のところで解説しました、その中で説明した図面値－実測値＝残差で考えます。

そうすれば標準偏差が計算できます。

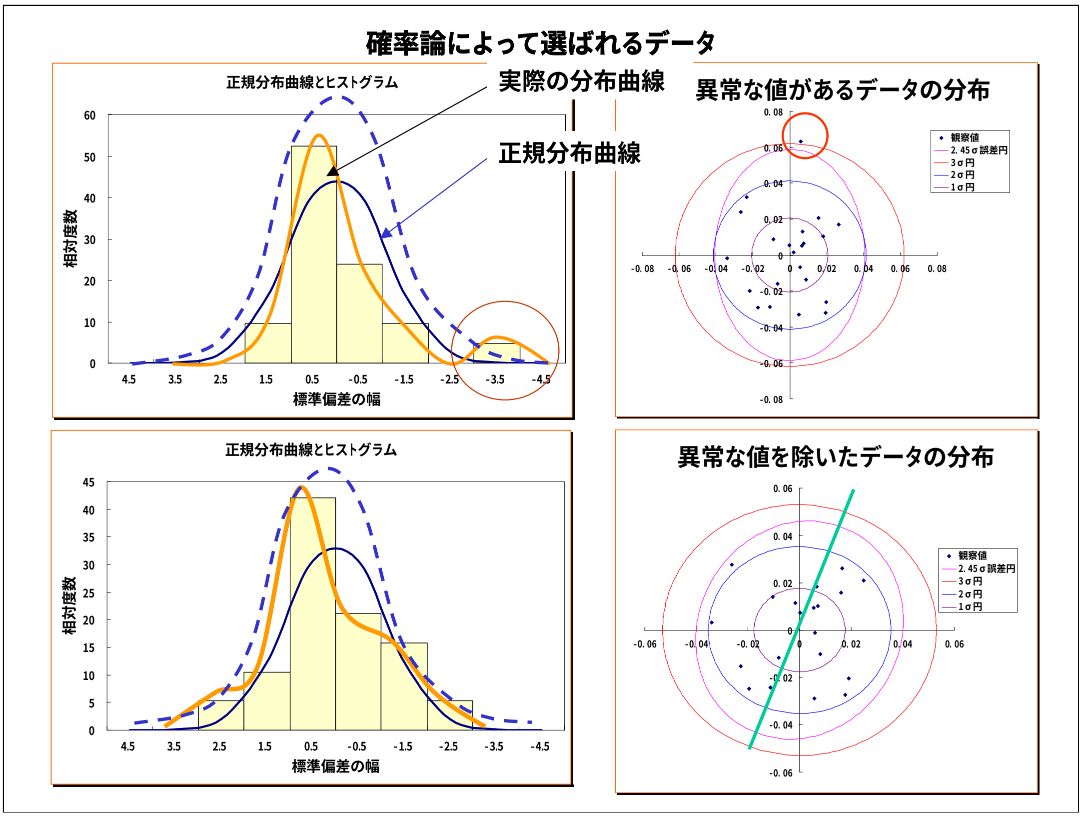
図面データも現地データも実測値でこの二つの比較をする訳です、一般的には現地データが図面データより正確であることを前提に話していきます。



座標位置の誤差の場合は①与点が同じ場合と②与点又は座標系が異なる場合があります。

それと測量の場合は測量器機、測量方法、データの補正方法が同じでないという要因があります。

図面が古いほど図面には様々な問題や誤差を大きくしている要因が含まれています、これらは計算以前に調べて排除できるものもあります、それについては復元事例を参考にして頂くとしてここでは数理の扱いを主に解説します。



基本的にある一定の条件の基に得られる結果と言いますかデータの差は正規分布に従うことが判っています。

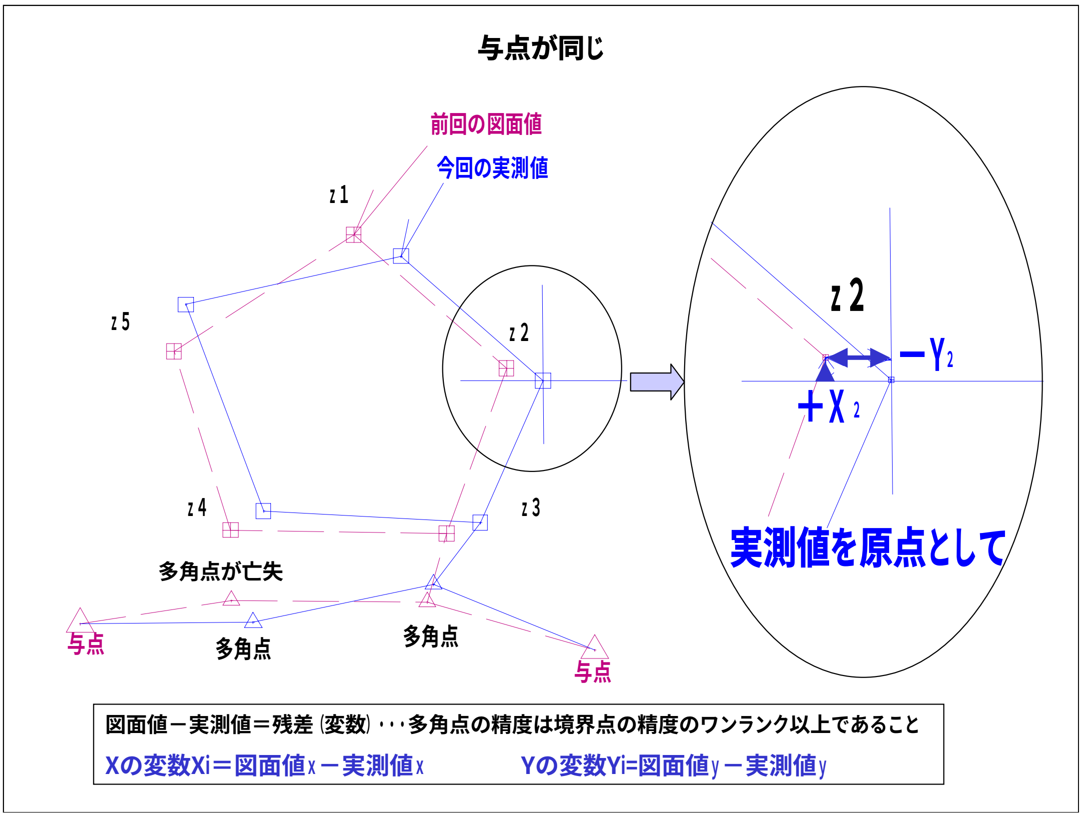
要するに得られた幾つかのデータの中から正規分布に従うデータの組み合わせを探せばその図面値と現地の値が本来持っている本来の性能を表していると考えられる訳です。

実際のデータは正規分布の確率に完全に収まることはありません、ある程度の余裕を持ってその中に収まるかどうかの判断を下すわけです、これを図のような有意水準とイメージとして貰えればよいと思います。

・・・難しいことは置きまして、このような状態を統計的な手法で選ぶことです確率論とは、この手法は統計学ではすでに考えられていますので我々はそれを利用するだけです。

詳しくはファイル「χ^2検定・ｔ検定」をご覧ください。

実際の検定はExcel「henkan.　xls」でｔ異常値検定とχ^2正規分布適合度検定　で使っていますのでそちらをご利用ください。



与点があり、多角点が亡失しているときに多角点を新設して測量をし、境界点を測った結果です。

過去の図面値と現在の実測値の相対位置誤差の計算になります。

図面値と実測値があり、それぞれXとYがありますのでこのどちらかを基準にした変数を考えます。

通常実測値が正確な事が多いので変数は図面値−実測値＝残差で残差を変数とします。

Xの変数Xi＝図面値ｘ−実測値ｘ、Yの変数Yi=図面値ｙ−実測値ｙです。

注意点は多角点の精度は境界点の精度のワンランク上であることが必要です、多角点の精度が低い場合は次のスライドへ

与点が同じ二変数標準偏差の計算

№	図面値			実測値			変数Xi	変数Yi
	点名	Xa	Ya	点名	Xb	Yb	Xb-Xa	Yb-Ya
1	z1	98132.033	-12812.170	g1	98132.046	-12812.184	0.013	-0.014
2	z2	98074.220	-12792.521	g2	98074.220	-12792.518	0.000	0.003
3	z3	98079.396	-12798.465	g3	98079.379	-12798.469	-0.017	-0.004
4	z4	98084.969	-12798.969	g4	98084.964	-12798.955	-0.005	0.014
5	z5	98093.863	-12798.901	g5	98093.857	-12798.889	-0.006	0.012
6	z6	98095.638	-12798.189	g6	98095.627	-12798.188	-0.011	0.001
7	z7	98115.828	-12798.162	g7	98115.812	-12798.155	-0.016	0.007
8	z8	98147.233	-12796.223	g8	98147.234	-12796.233	0.001	-0.010
9	z9	98137.638	-12833.485	g9	98137.608	-12833.489	-0.030	-0.004
10	z10	98131.510	-12836.571	g10	98131.511	-12836.576	0.001	-0.005
11	z11	98131.144	-12839.798	g11	98131.127	-12839.794	-0.017	0.004
12	z12	98131.161	-12839.948	g12	98131.142	-12839.944	-0.019	0.004
13	z13	98131.289	-12841.053	g13	98131.290	-12841.059	0.001	-0.006
14	z14	98131.768	-12845.202	g14	98131.762	-12845.201	-0.006	0.001
						平均	-0.0079	0.0002
						標準偏差	0.0111	0.0079
						標準偏差	0.0097	

数	14

エクセルの関数 **STDEV** $(X_1:Xn)$

次の式でも求まる

標準偏差 $\sigma x=\sqrt{(\Sigma((Xi-Xi平均)^2/(n-1))}$

標準偏差 $\sigma y=\sqrt{(\Sigma((Yi-Yi平均)^2/(n-1))}$

二変数（二次元）の標準偏差

$$\sigma=\sqrt{(\sigma x^2/2+\sigma y^2/2)}$$

$$\sqrt{(0.0111^2/2+0.0079^2/2)}=0.0097$$

左に図面値、右に実測値で変数xiと変数yiを計算します。

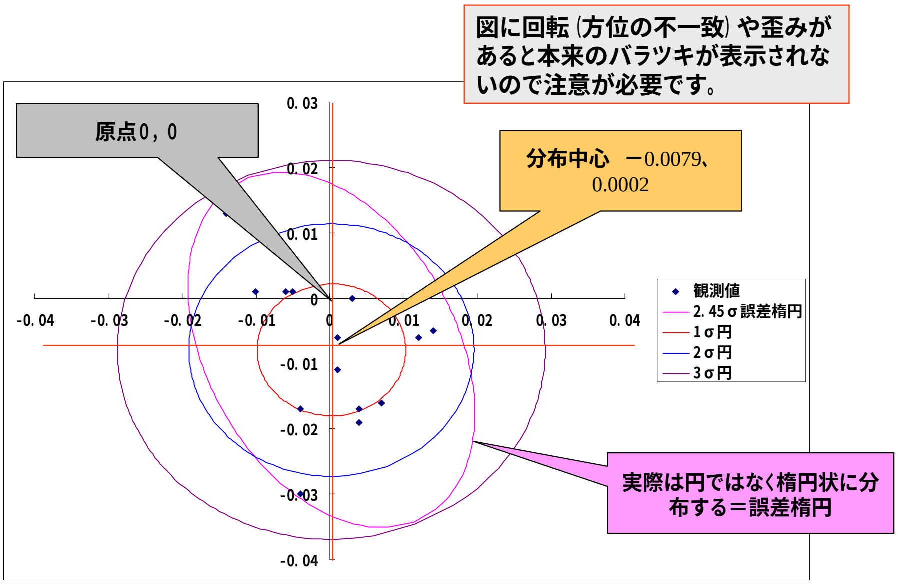
セクセルの関数　STDEV　を使ってxiの標準偏差0.0111、yiの標準偏差0.0079を求めます、ここまでは一変数の標準偏差の計算と同じです。

したの枠に式は相関関係のない二変数の標準偏差の計算式です、この式から二変数の標準偏差0.0097を求めます。

これを散布図を書いて図でその関係を確認してみます。

一変数の標準偏差の計算は計算式で求めてもExcel関数STDEVで求めても同じ数字になります。

与点が同じ散布図



x、yの平均値を中心に分布する。

散布図の分布は円ではありません、紫の楕円のとおり実際の分布は楕円になります。

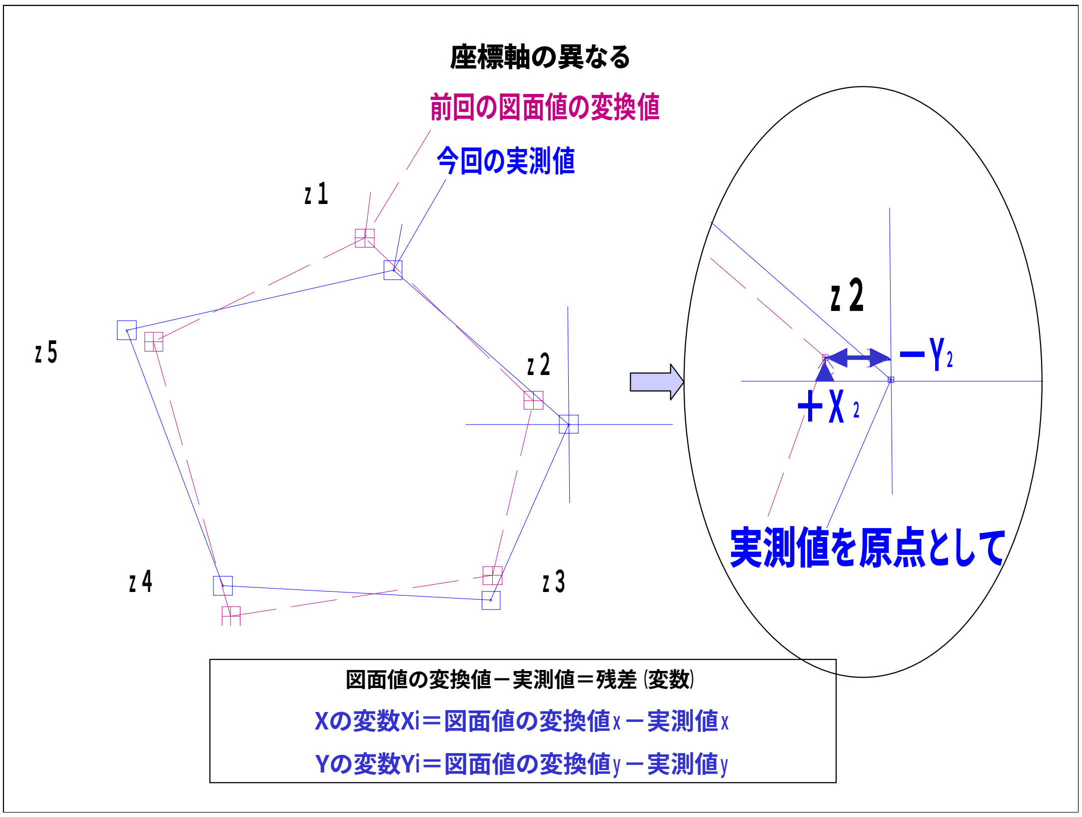
精度が高いほど円に近づきます。

分布の中心点は原点、つまりx＝0，y＝0ではなくxの平均値、yの平均値の交点を中心とした楕円状の分布します。

後で最小二乗法よって計算した散布図と標準偏差が出てきますが最小二乗法による場合は分布の中心点は原点、つまりx＝0，y＝0、x平均値＝0、yの平均値＝0となります。

また、標準偏差も0.0097から0.0091と小さくなります、これは与点が多少回転していることが原因です。

この例は後で説明します。



座標系が異なる、与点が異なる、たとえば任意と世界測地、旧日本測地と世界測地、任意と任意などです。

与点に精度がない、測量器機が異なる、測量方法が異なる、データの補正方法が異なるなどでは単純に比較できないデータの場合は実測値ベースで座標変換して重ねて比較することになります。

座標変換の方法は最小二乗法により、図面値の変換値と実測値があり、それぞれXとYがありますのでこのどちらかを基準にした変数を考えます。

通常実測値が正確な事が多いので変数は図面値の変換値−実測値＝残差で残差を変数とします。

Xの変数Xi＝図面値の変換値ｘ－実測値ｘ、Yの変数Yi=図面値の変換値ｙ－実測値ｙです。

注意点は図面の伸縮、歪み、座標軸の回転などに注意して変換することが要求されます。

座標軸の異なる二変数標準偏差の計算

最小二乗法による計算値

№	図[illegible]の変換値			実測値			変数Xi	変数Yi
	点名	Xa	Ya	点名	Xb	Yb	Xb-Xa	Yb-Ya
1	Hz1	98132.026	-12812.172	g1	98132.046	-12812.184	0.020	-0.012
2	Hz2	98074.215	-12792.515	g2	98074.220	-12792.518	0.005	-0.003
3	Hz3	98079.391	-12798.460	g3	98079.379	-12798.469	-0.012	-0.009
4	Hz4	98084.964	-12798.964	g4	98084.964	-12798.955	0.000	0.009
5	Hz5	98093.858	-12798.898	g5	98093.857	-12798.889	-0.001	0.009
6	Hz6	98095.633	-12798.186	g6	98095.627	-12798.188	-0.006	-0.002
7	Hz7	98115.823	-12798.162	g7	98115.812	-12798.155	-0.011	0.007
8	Hz8	98147.228	-12796.227	g8	98147.234	-12796.233	0.006	-0.006
9	Hz9	98137.628	-12833.488	g9	98137.608	-12833.489	-0.020	-0.001
10	Hz10	98131.499	-12836.573	g10	98131.511	-12836.576	0.012	-0.003
11	Hz11	98131.133	-12839.800	g11	98131.127	-12839.794	-0.006	0.006
12	Hz12	98131.150	-12839.950	g12	98131.142	-12839.944	-0.008	0.006
13	Hz13	98131.278	-12841.055	g13	98131.290	-12841.059	0.012	-0.004
14	Hz14	98131.756	-12845.204	g14	98131.762	-12845.201	0.006	0.003
						平均	0.0000	0.0000
						標準偏差x	0.0110	0.0068
						標準偏差	0.0091	

数	14

（注意） 平均が0になる

与点が同じの標準偏差は0.0097で変換値が0.0091なのは図が回転しているからです。

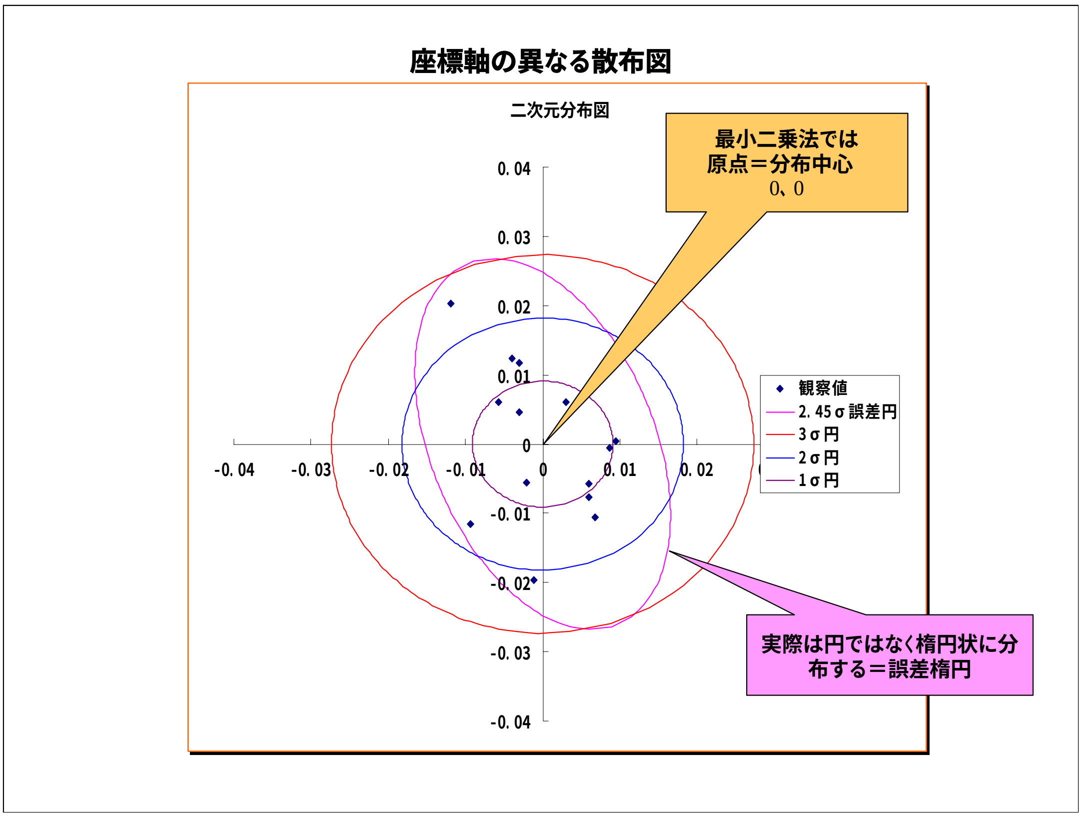
標準偏差が「与点が同じ」と違うのは図が回転しているからです。

最小二乗法による座標変換を使いますので平均値はx、yとも0になります。

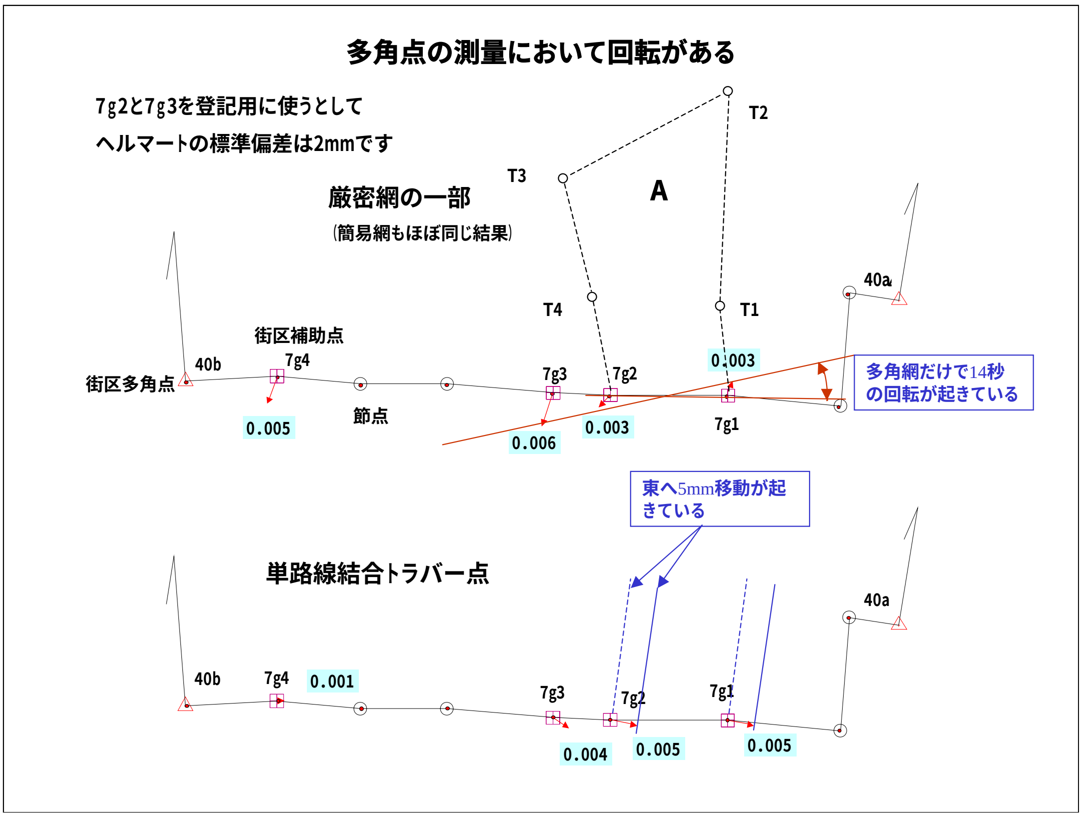
図の回転も修正されますので「与点が同じ」の標準偏差より若干小さくなります。

この例では0.0097から0.0091に小さくなっています。

これを散布図で見てみますと。



分布の中心が原点x=0, y=0、平均値x=0、y=0となっています。
楕円の形は先の「与点が同じ」と変わりません。



図の回転、移動は多角測量によっても起こります。

この図は最近の街区多角点を使って厳密網と通常の単路線結合トラバーの結果を示しています、同じデータで計算方法が違います。

三角形が街区多角点、□が街区補助点、観測誤差のバランスを取るため○の節点を追加して測量しています。

観測結果を厳密網と通常の結合トラバーで計算した結果をベクトル図で見たものです。

上の図をみると7g1と7g2に相対的な回転が14秒あるのでAの結果に回転が残り正しい標準偏差の計算はできなことになります。

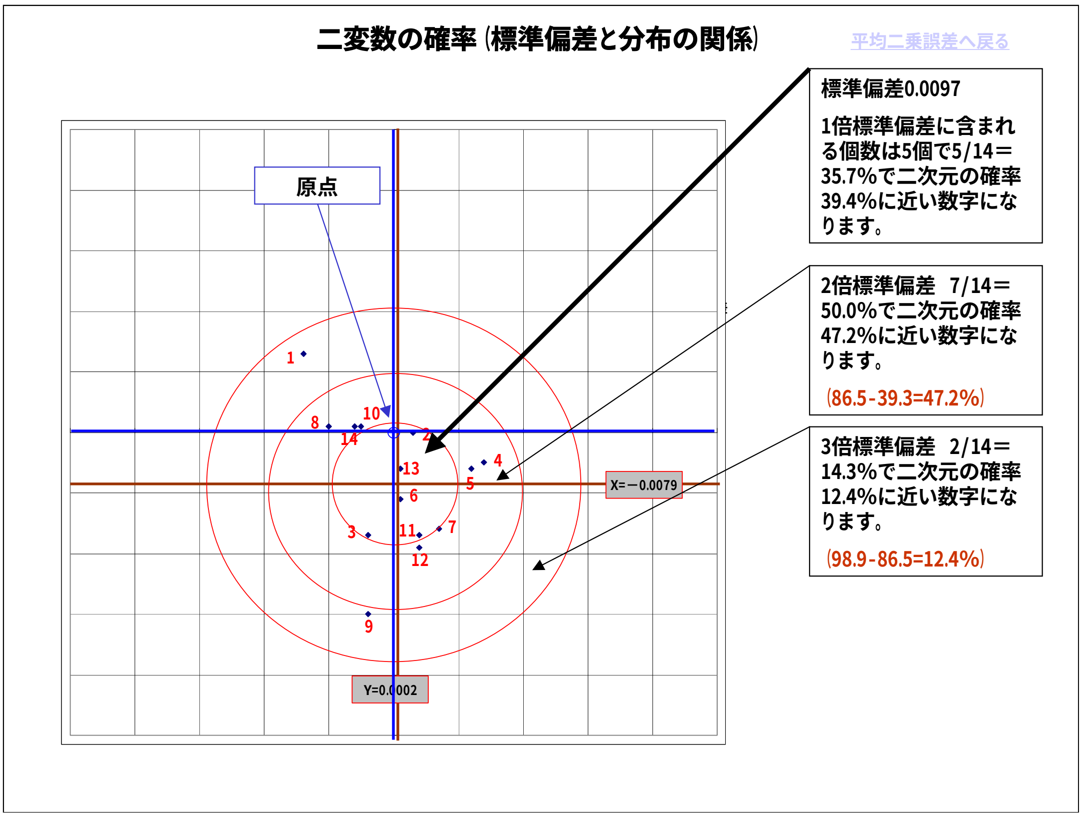
下の図は普通に使われている結合トラバーの結果です、この場合は移動がみられるので結果的にAの土地も西へ全体が5mm移動した結果が得られる。

この場合標準偏差をそのままのデータで計算しても平行移動による場合は標準偏差の結果に与える影響は少ない。

実際にはこのような解析はしませんので回転と移動は当然起きている、また測距の道具が違えば伸縮も起きているのです。

このような事が起きている前提で考えれば常に計算処理したデータで標準偏差を考える事が必要です。

いかに正確な測量をしても使う多角点に回転や移動がるということです。



散布図でその標準偏差の円の中にある点の割合を見てみます。

先ほど計算したXiとYiの平均値を赤の直線で入れてみます、この直線の交点を基に先ほど計算した標準偏差の1倍、2倍、3倍の円を入れてみます。

これが正規分布と言われるものです、この円の中にあるデータが確率というものです。

この図からデータは原点を中心にしてるのではなく平均値を中心にあるということです、別のところで平均二乗誤差の説明をしますが標準偏差と平均二乗誤差の違いはここにあります。

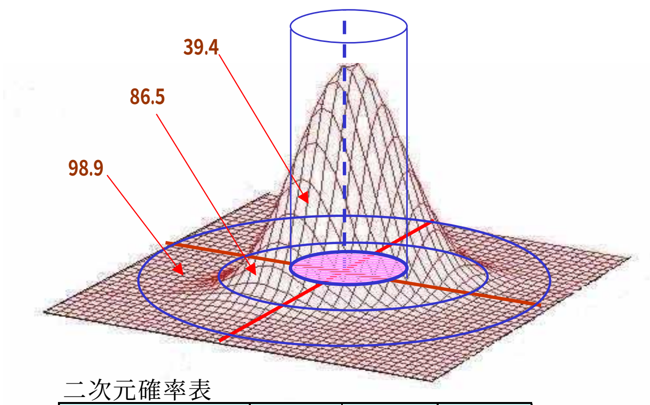
データ数が少ないので１標準偏差内では39.3%に対して35.7%と一致しませんがデータ数が千個単位に成ると期待される確率に近づきます。

期待する確率の数値は次のスライドにあります、ここで注意していただきたいのは我々が普通耳にする確率はプラスマイナス1標準偏差で68.3%ですがこれは変数が1つの場合の確率であって座標値の場合は変数が2つの二次元正規分布になり1標準偏差の円の中にある確率は39.3%です。

二変数　正規分布図と確率

二次元正規分布図

（散布図からイメージ）



二次元確率表

標準偏差	1σ	2	3
確率（約）	39.4	86.5	98.9

標準偏差	確率		
0.05	0.0012	2.05	0.8777
0.1	0.0050	2.1	0.8897
0.15	0.0112	2.15	0.9009
0.2	0.0198	2.2	0.9111
0.25	0.0308	2.25	0.9204
0.3	0.0440	2.3	0.9290
0.35	0.0594	2.35	0.9368
0.4	0.0769	2.4	0.9439
0.45	0.0963	2.45	0.9503
0.5	0.1175	2.5	0.9561
0.55	0.1404	2.55	0.9613
0.6	0.1647	2.6	0.9660
0.65	0.1904	2.65	0.9701
0.7	0.2173	2.7	0.9739
0.75	0.2452	2.75	0.9772
0.8	0.2739	2.8	0.9802
0.85	0.3032	2.85	0.9828
0.9	0.3330	2.9	0.9851
0.95	0.3632	2.95	0.9871
1	0.3935	3	0.9889
1.05	0.4238	3.05	0.9905
1.1	0.4539	3.1	0.9918
1.15	0.4838	3.15	0.9930
1.2	0.5132	3.2	0.9940
1.25	0.5422	3.25	0.9949
1.3	0.5704	3.3	0.9957
1.35	0.5980	3.35	0.9963
1.4	0.6247	3.4	0.9969
1.45	0.6505	3.45	0.9974
1.5	0.6753	3.5	0.9978
1.55	0.6992	3.55	0.9982
1.6	0.7220	3.6	0.9985
1.65	0.7437	3.65	0.9987
1.7	0.7643	3.7	0.9989
1.75	0.7837	3.75	0.9991
1.8	0.8021	3.8	0.9993
1.85	0.8194	3.85	0.9994
1.9	0.8355	3.9	0.9995
1.95	0.8506	3.95	0.9996
2	0.8647	4	0.9997

正規分布について復習しておきますと

一次元正規分布図と言われるもので皆さんは日調連の測量実務要領、福永先生の14条地図活用マニュアルとかに出てきますのでご存じのものです。

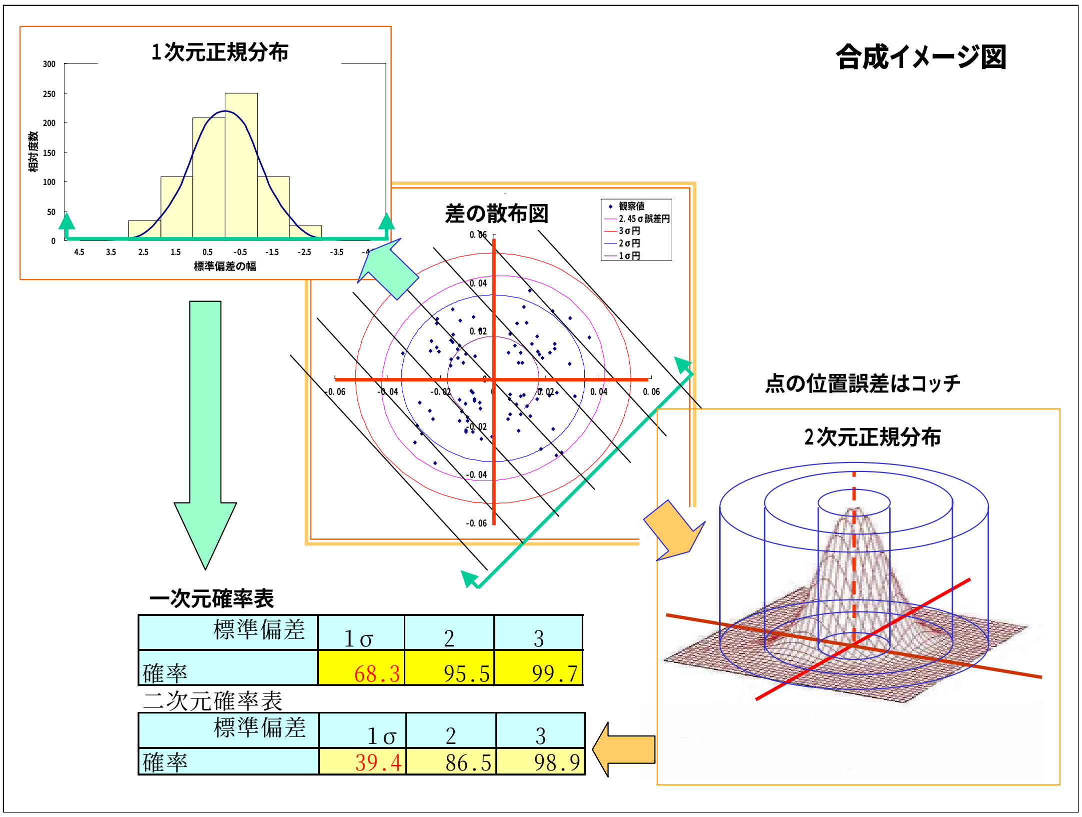
一次元正規分布図は変数、つまりデータの種類が一つの場合の分布です、測量で言えば距離だけ、角度だけの場合です。

ここでは二次元正規分布図で因数が二つ、座標値の場合はxとyの二つの変数からなりますのでこの分布になります、測量図と言われる点の表示は全て二次元正規分布になります、ですから確率が一次元正規分布とは異なります。

詳しい数値は右の表にありますので参考にしてください。

測量学といわれる専門書はこれと同じ解説になっていますが測量実務要領、福永先生の14条地図活用マニュアルとか地籍調査には誤った解説がなされていることがありますので注意してください。

ここでの説明とは関係ありませんが標準偏差＝平均二乗誤差とう解説も偶に見かけますが標準偏差と平均二乗誤差は全く違った指標ですから注意してください。

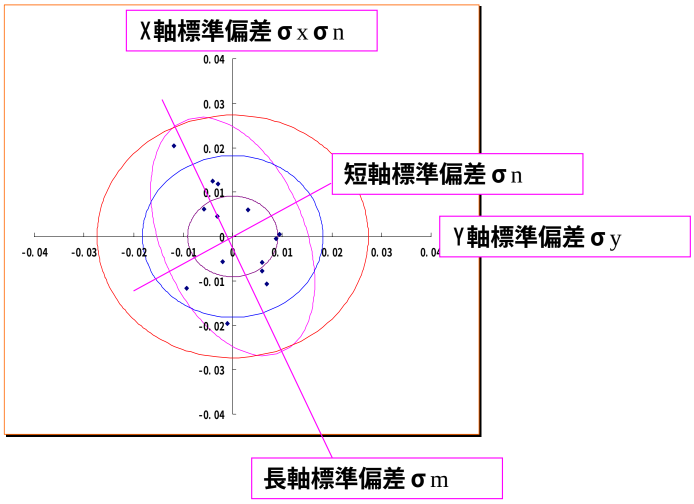


標準偏差	1σ	2	3
確率	68.3	95.5	99.7

標準偏差	1σ	2	3
確率	39.4	86.5	98.9

散布図から二変数、二次元正規分布をイメージしたのが右下の絵です。
左上は45軸の値を一変数、一次元正規分布に当てはめて見た例です。
確率の関係も表にしてみました。

二変数　標準偏差の関係



$$\sigma = \sqrt{\sigma x^2/2 + \sigma y^2/2}$$
$$= \sqrt{\sigma m^2/2 + \sigma n^2/2}$$

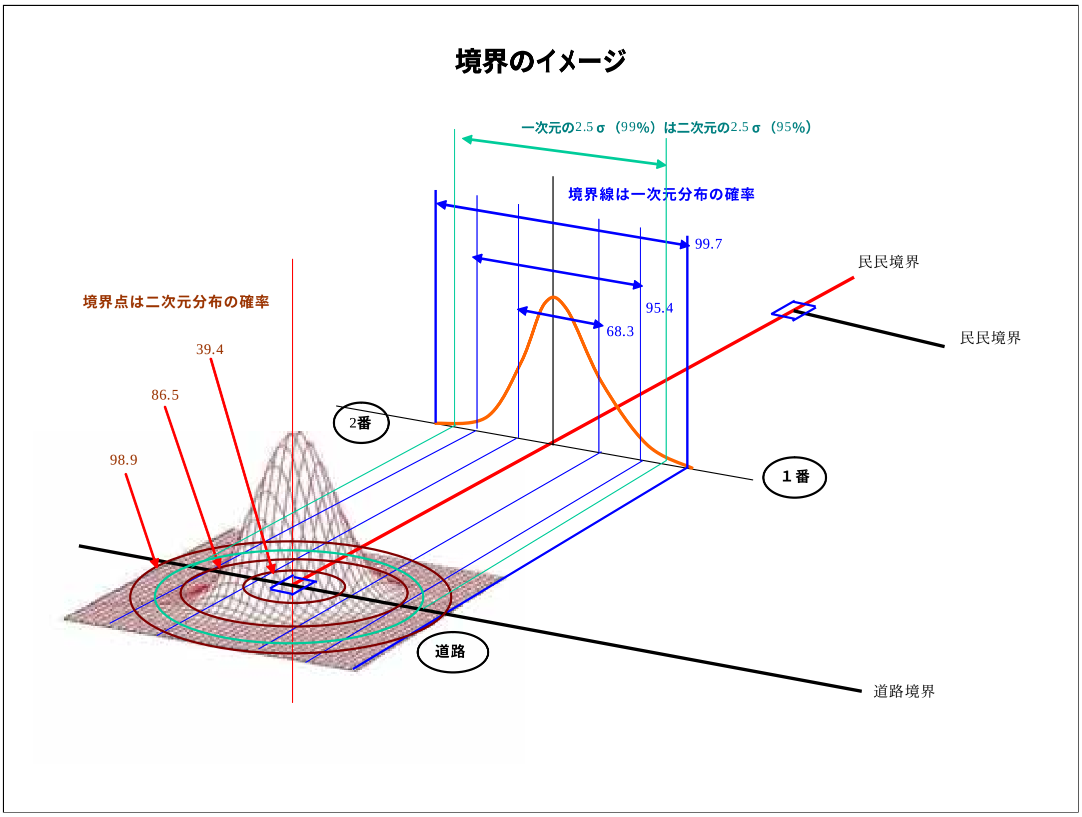
二変数の標準偏差の計算式は表の通りです、この標準偏差は誤差楕円の長軸の標準偏差と短軸の標準偏差を使っても同じ結果になります。

別の言い方では直交する任意の軸では同じ結果が得られます。

私も暇に任せて様々な計算をしてみましたがその通りでした、確率された学問とはこのようなものなのでしょう。

その図面の精度は基本的に標準偏差で判断しますが実際にはデータ解析を経て最終的に判断することになります、そのことは事例集を見て頂き勉強するしかありません。

解った振りをして独学するなと言いたいのですが独学すらしない土地家屋調査士が多いのでなんともはや！　です。



せっかくここまで話をして二変数の分布と境界標の関係をイメージ出来ないともったいないので簡単に図で説明します。

境界点のイメージは二次元正規分布、線のイメージは一次元正規分布ですが我々が考えるのは点の位置誤差ですからあくまでも二次元の正規分布ということになります。

しかも楕円なのですが楕円でやっていきますとかなりと言いますかめちゃくちゃというほど難しくなりますので円として多くは説明してあります。