2011/12/12

座標解析と最小二乗法による境界復元

公差と不良率

国土調査法、不動産登記法に定められている筆界の位置誤差は単純に不良率0．01％以下になっているか否かを確認すれば済むことです。
不良率を計算する技術がないために平均二乗誤差やら筆界点間の辺長差の基準が設けられている、これが判りにくくしている大きな要因である。
不良率は単に平均二乗誤差や標準偏差で計算できるものでは無く、相関係数、二変量の標準偏差、相関係数、分布のずれ（⊿x、⊿y）から計算されるものである。
ここでは不良率を管理する場合の一変量のデータとの変量のデータについて解説した。座標の位置誤差は二変量なので一変量は関係ないが基礎知識として必要なので始めに説明してみた。
Henkan．xls プログラム（Book）では不良率計算ができるようにプログラムされている。

土地家屋調査士・測量士 小野孝治　　http://012.o.oo7.jp/

「標準偏差と平均二乗誤差の違い」を先に読んでください。

いかに二変量で有ろうとも成果（図面値）と実測値（検査値）があれば不良率管理でそのデータ良否が簡単に判定出来ます。

一変量の場合は標準偏差、公差、公差中心と分布中心のズレから不良率が求まりますので不良率の限度を決めておけば「良」「不良」の判断ができます。

二変量の場合は標準偏差、公差、公差中心と分布中心のズレ、相関係数から不良率が求まりますので不良率の限度を決めておけば「良」「不良」の判断ができます。

二変量の場合は計算が複雑になるため今まではあまりこのことを言う方がおりませんでした、そこで国土調査では平均二乗誤差という指標でもって判断しています。

「標準偏差と平均二乗誤差の違い」にも解説して有りますが不良率の判定指標として平均二乗誤差は適切でないと言わざるを得ません。

コンピュータの発達した現代ではソフトが組めれば標準偏差、公差、公差中心と分布中心のズレと相関係数から不良率を求めることが簡単にできますのでこの方法に切り替えるべきと考えます。

公差と確率・不良率について

製品の全部について検査をすることはコストの面で出来ません、そこで一部を抽出して検査しその結果から製品全体の良、不良を判断します。

一変量の公差と確率・不良率について
　境界は二変量のデータを扱いますので二変量の解説をしたいのですがまずは簡単な一変量で解説します。

　身近な例からJIS１級鋼製巻尺について解説します、鋼製巻尺JIS１級許容差50m±5.2mmとなっています。
　これを1000本作って販売するとします、原価5,000円で販売価格8,000円で計画します、一本当たりの検査費用が6,000円かかります。
　これでは11,000になり3,000円／本の赤字になりますので100本の抜き取り検査をして販売することにします。
　100本＊6,000円／1,000本＝600円／本　で原価＋600円で5,600円となり2,400円／本のもうけが期待出来ます。

　このときの抜き取り検査の結果を公差と確率・不良率を使って判定します。

一変量のデータと二変量のデータの基本的な考えは同じなので基礎知識として覚える必要が有ります、それをベースに二変量のデータを覚える手順になります。

境界の位置は二変量のデータを扱いますので二変量の解説をしたいのですがまずは簡単な一変量で解説します。

一変量のデータ例として身近に使われている鋼巻尺の製造について考えて見ました。

製品の全部について検査をすることはコストの面で出来ません、そこで全体の一部を抽出して検査しその結果から製品全体の良、不良を判断します。

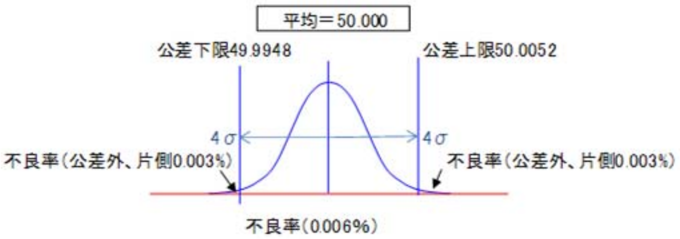
なぜこのような方法をとるのかといえばコストの低減と納期の短縮が出来るからです。

公差と標準偏差

公差は50m±5.2mmです。
　不良が発生しないように製品のバラツキをコントロールしますので確率・不良率を設定します。
　一般的には不良率が0.01％(1/10,000)以下になるように製品のバラツキをコントロールします、ここで使うのが標準偏差と確率です。

標準偏差	確率
3.89	0.9998998
3.9	0.9999038
3.91	0.9999077
3.92	0.9999115
3.93	0.9999151
3.94	0.9999185
3.95	0.9999218
3.96	0.9999251
3.97	0.9999281
3.98	0.9999311
3.99	0.9999339
4	0.9999367
4.01	0.9999393
4.02	0.9999418
4.03	0.9999442

10,000分の1以下になる確率0.9999以上は左の表から3.9倍標準偏差以上なので切りのよい4.0倍標準偏差(確率0.99994)を使います。
そこで公差5.2mmの4分の1＝1.30mm以下を標準偏差の目標とすればよく、このときの不良率は0.006%になります。



1,000本のスチールテープが標準偏差1.30以下で平均値が公差の中心の50.000の場合不良率が0.006%以下になることを目標に作ります。

製品精度のコントロールには平均値とバラツキの二つです、バラツキは標準偏差であらわしますので 標準偏差と確率・不良率の関係が重要になってきます。

ここでは標準偏差と確率・不良率の関係を理解出来ればいいでしょう。

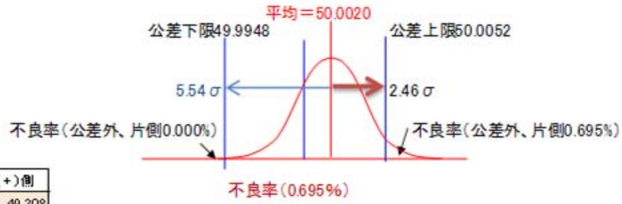
ここでは不良率0.01％以下にするということで4倍標準偏差（4σ、不良率0.006％）で設定しています。

マイナス側の不良率0.003％、プラス側の不良率0.003％で全体で0.006％になります。

物作りの現場では4σは普通の目標値で決して厳しい数値では有りません。

平均値のズレと不良率

1,000本の中からランダムに抜き取って100本のデータを調べたところ標準偏差1.30、平均値が50.0020だった。



標準偏差　σ	1.3		
公差の中心値	0		
公差の幅	5.2		
分布の平均値	2	(−側)	(+)側
全体確率　%	99.308	50.000	49.308
全体異常率　%	0.692	0.000	0.692

不良率の目標0.01％を越えてしまいます。
この場合全体が不良であると判断します。

処置
製品に対してはJIS一級ではなくJIS二級（±10mm）の製品として販売する・・価格は安くなる。

製造
平均値の狙い値を50.000に近づける・・・50.0010とかに狙値を変える。
バラツキの少ない製造方法に変える・・・標準偏差を1.06mm以下にする。

生産された鋼巻尺を検査しますと平均値が公差の中心に来ませんので公差の中心がずれた分だけ標準偏差を小さくコントロールする必要があります。

平均値がある程度バラついても不良が出ないようにするには標準偏差をコントロールするのが安全です、実際には4σではなくもっと安全な数値でコントロールするはずです。

マイナス側とプラス側で平均値がプラスに出ますとその分だけ標準偏差を小さくしないと不良率が基準を超えてしまいます。

分布図は平均値0.0020だけプラス側に移動しますと公差を超える量が多くなります、不良率0.003％が0.695％になり結果全体が不良とされます。

これを解消するには標準偏差1.3mmでは駄目で1.06mm程度まで厳しくコントロールする必要が出てきます。

例題1

1,00本の中からランダムに抜き取って10本のデータを調べたところ次のデータが得られた、
不良率の目標0.01%以下として100本全体への評価は 良品 か 不良品か。
10本の製品は全て公差内(5.2mm)であった。

	テープ検定
1	50.0021
2	50.0045
3	50.0011
4	50.0032
5	50.0047
6	50.0035
7	50.0028
8	50.0043
9	50.0050
10	50.0026
平均	50.0034
標準偏差	0.00126

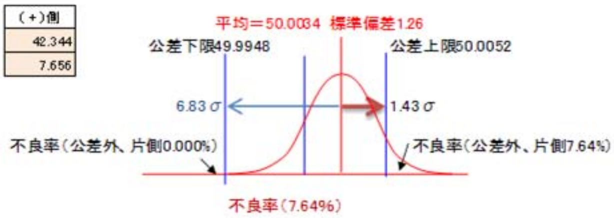
今までの考えで計算してみましょう。

一個一個は公差内だけれど製品全体としての判断はどうかという問題です。

回答1

標準偏差、公差中心から分布中心のズレ(公差中心値ー分布中心値)を計算すれば不良率は求まります。
HenkanV1.2以降のシート「huryou」に計算プログラム(Book)が有りますので参考にしてください。

		(-側)	(+)側
標準偏差 σ	1.26		
公差の中心値	0		
公差の幅	5.2		
分布の平均値	3.4		
全体確率 %	92.344	50.000	42.344
全体異常率 %	7.656	0.000	7.656



不良率の目標0.01%を越えてしまいます。
この場合全体が不良であると判断します。
ここで重要なのは抜き取った製品が公差内であっても全体としては【不良】とされることです。

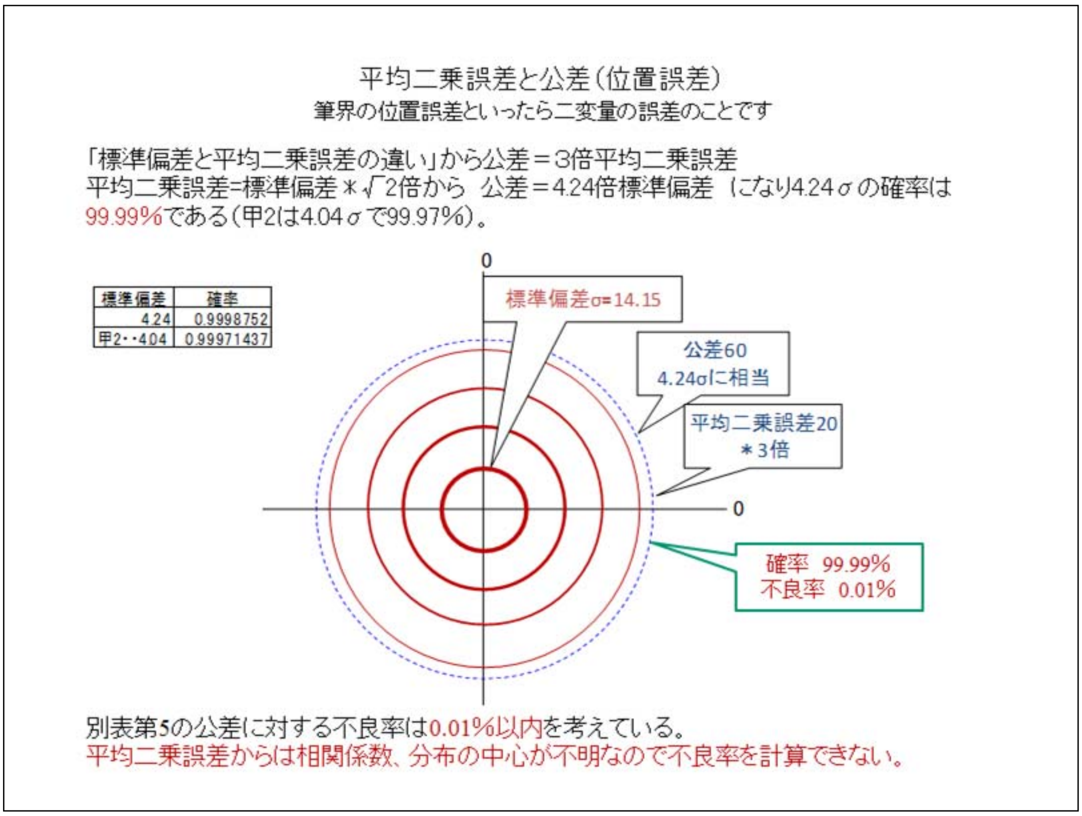
標準偏差、公差中心から分布中心のズレ(公差中心値-分布中心値)を計算すれば不良率は求まります。

エクセルの関数で求まりますのでなんら難しいことはありません。

標準偏差1.26、中心のズレ3.4、公差5.2の3要素から計算できます、マイナス側の不良率はほぼ0、プラス側が平均値がプラスに移動した分だけ不良率が7.66%と大きくなります。

結果、抜き取った製品が公差内であっても全体(母集団)としては【不良】とされます。

この辺が皆さん以外と判っていないところです。



位置誤差はx、yの二変量で表します、位置誤差の公差といえば二変量、つまりx、yが出てきますので二変量の分布になります。

公差は約3倍平均二乗誤差に設定されています、平均二乗誤差と標準偏差の関係は公差＝4.24倍標準偏差になります、4.24σの確率は99.99％である（甲2は4.04σで99.97％）です。

このことから不良率は0.01％以下を想定して別表第5の公差と平均二乗誤差がきめられていると計算できます。

「標準偏差と平均二乗誤差の違い」から標準偏差と平均二乗誤差と公差の関係を復習してみました。

二変量の場合、相関係数と公差中心と分布中心のズレが実際にはあり平均二乗誤差ではこれを考慮して計算できないので標準偏差で計算する必要があります。

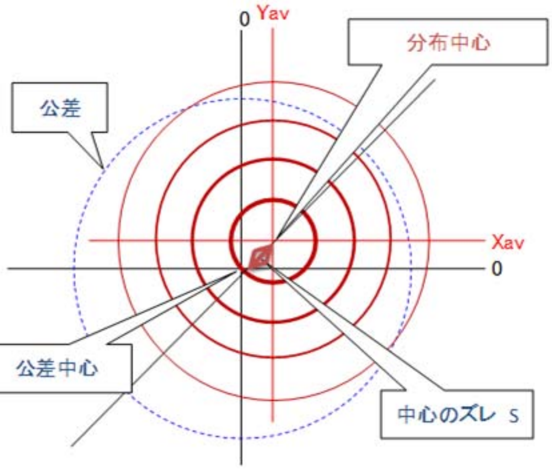
この図は公差60で標準偏差は60/4.24＝14.2、xとyのズレが0、相関係数が0で不良率0.01％以下の状態です、平均二乗誤差は20です。

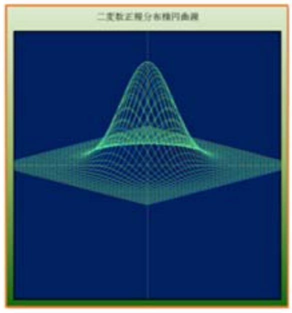
確率は公差内にある密度の合計です、不良率は公差外にある密度の合計ですが公差外の密度は無限に広がるため公差内の密度を求めて密度の総計から確率を引いたものが不良率です。

二変量に於ける中心のズレと不良率

別表第5の位置誤差の不良率を0.01％以内として標準偏差、相関係数、分布中心距離、公差から不良率を計算して筆界の良・不良の判断をする。
（甲2は0.03％であるが0.01％で考える、平均二乗誤差に換算すれば67mmである）

二変量には分布中心のズレによって公差内の確率は低くなる（不良率が大きくなる）。





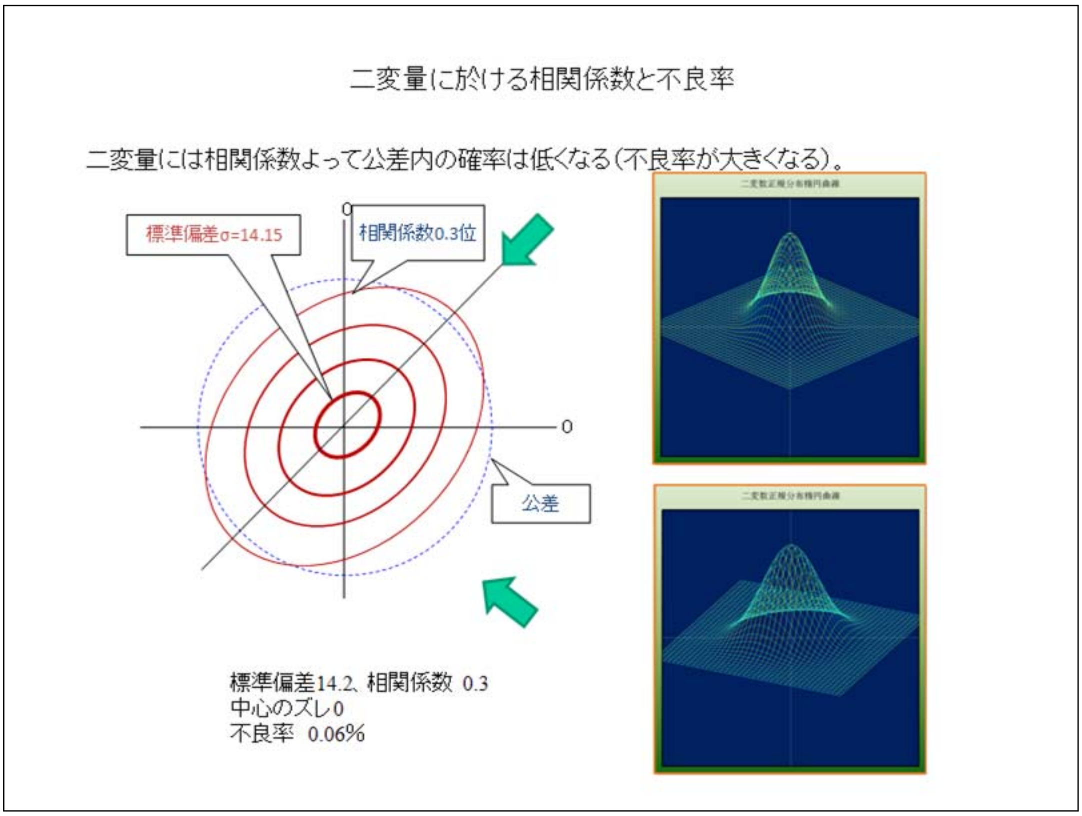
標準偏差14.2、相関係数 0
中心のズレ14
不良率 0.11％

甲2の基準が平均二乗誤差70mm、公差200mmで不良率が0．01％が0.03％と違いますがこれは200mmから平均二乗誤差を逆算すると67mmとなり切り上げて70mmとしたと考えられます。

同じ標準偏差でも公差中心から分布中心のズレが多くると確率はさがります。

この図では不良率が0.01％（中心のズレが0）から中心のズレが14mmあれば不良率は0.11％と大きくなります。

右に立体的な分布を示しました。

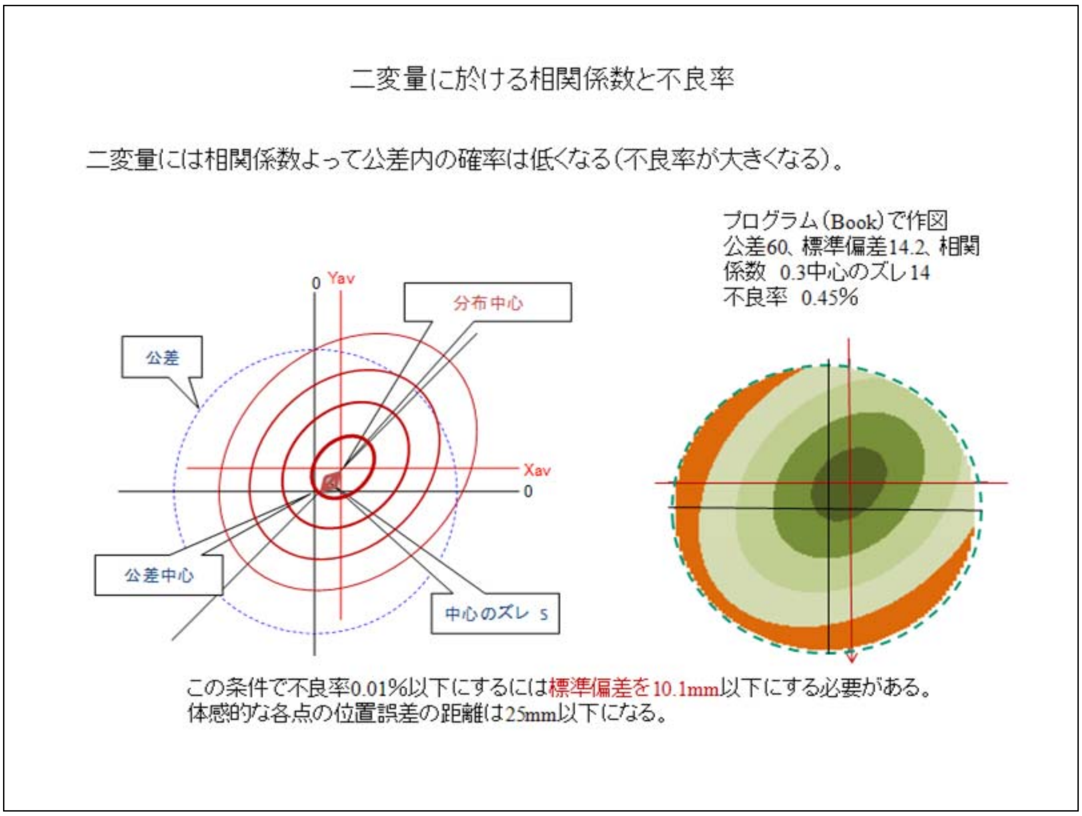


二変量では分布が楕円に成ります、この状態を相関係数を使って表します、実際のデータでは相関係数が0と言うことはあり得ません。

同じ標準偏差でも相関係数の絶対値が大きくなると不良率が上がります。

この図では不良率が0.01%(中心のズレが0)から相関係数が0.3になったとき不良率は0.06%と大きくなった状態の図です。

右の図は緑の矢印からみた立体的分布です。



実際には公差中心から分布中心のズレと分布が楕円（相関係数が0より大きい状態）の状態にあります。

これは公差60に対する不良率を計算しています、平均二乗誤差では相関係数、中心のズレがあると不良率が計算できないので標準偏差で計算することになります。

相関係数0.3と中心のズレ14mmが同時に起きた場合（通常はこのような条件で分布する）不良率は0.01％から0.45％と大きくなります。

この状態で0.01％以下の不良率にするには標準偏差を10.1mm以下にしなければなりません、平均二乗誤差に換算（おおよその値）すると14.2mmです。

相関係数、公差中心から分布中心のズレは基準点の影響を強く受けますので不良率が高い場合は標準偏差を小さくするしか有りません。

左の図はイメージですが右の図は実際にプログラム（Book）を使って計算した図です、だいたいイメージ通りですね。

別表5の位置誤差の公差は、例えば甲2では200mmとなっておりそれ相応に緩やかな基準と思われがちですが中心のズレと相関係数を加味すればそれほど緩い基準では無いことに気がつきます。

その他の公差

相関係数、分布中心のズレが考慮されていない数値

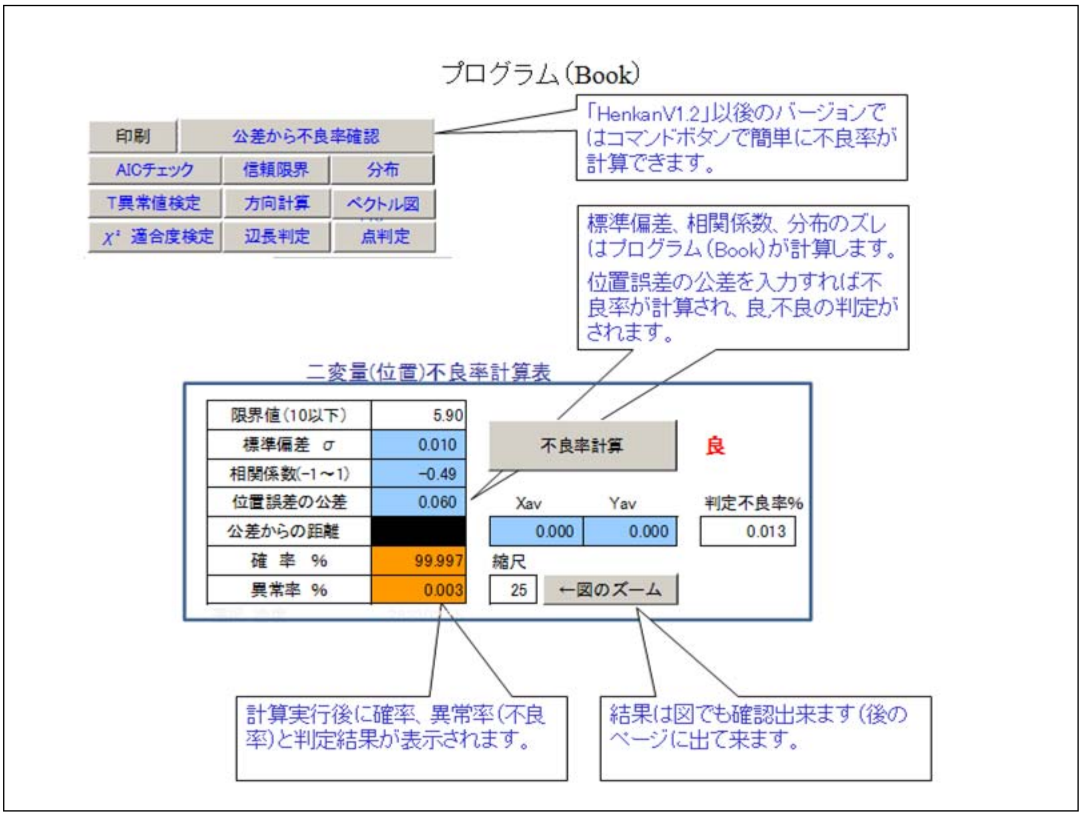
精度区分	位置誤差(与点からの) 平均二乗誤差	位置誤差(与点からの) 公差	図上距離と直接測定による距離との差異の公差(A)	出合差 mm	(数値法) 標準偏差＝公差／4.25
区画整理		**42**	街区 30m 10mm ・30m以上 S/3000 筆界 20m 10mm ・20m以上 S/2000	**20**	**10.0**
DID		**42**		20	**10.0**
14条 法務局		**42** 60	平地 20m 10mm ・20m以上 S/2000 山地 20m 20mm ・20m以上 S/1000	**20** 30	**10.0** 14.0
甲1	20	60	0.020m＋0.003√sm	30	**14.0**
甲2	35 (70)	106 (200)	0.040m＋0.010√sm＋αmm	50	**25.0** (47)
甲3	64 (150)	191 (450)	0.080m＋0.020√sm＋αmm	90	**45.0** (106)
乙1	85 (250)	255 (750)	0.130m＋0.040√sm＋αmm	120	**60.0** (175)
土地改良			地籍調査の精度区分甲2、甲3、乙1に同じ)は図解法 **太字は計算値**		

不良率はどの場合でも0.01％以下というとで、図面の区域毎に公差が決めてあれば充分ですが残念ながらありません。

そこでこの表は公差の規定がない区画整理地区、DID地区、法務局作成の14条地図と公差が図解法で示されている甲2～乙1の公差を試算してみたものです。

法律によって基準が皆違いますので本来は統一すべきものでしょう。

とくに測量機器の発達した現代に置いては甲1～乙3などの細分化の必要ないと考えられますが取りあえずこの表の値を使って良不良の判定をしてみることにします。



計算するにはプログラム(Book)が必要ですから簡単なものを作って見ました。

最終ページにあるような図も表示されます。

計算に必要な標準偏差、相関係数、公差からの距離は「HenkanV1.2」以後のバージョンで取得できます。

同じ基準点を与点に測量した場合はヘルマート変換は使いませんのでシートmuhenから直接不良率を計算することになります。

基準点が使えない場合でヘルマート変換を使う場合伸縮率1.00000にして計算し、不良率を求める必要が有ります。

測距の機器がスチールテープ以前の場合は伸縮が有りますので伸縮率をフリーで計算する必要が有ります。

測角がアリダードなどで不正確な場合はデータによってはアフィン変換を使う必要もあります。

「HenkanV1.2」以後ではそのプログラム(Book)も用意して有ります。

別表5は相関係数を 0 として考えているようだ

相関係数と確率

相関係数

標準偏差	0	0.1	0.2	0.3	0.4	0.5
1.0σ	39.35%	39.20%	38.62%	37.64%	36.21%	34.27%
1.5σ	67.53	67.09	66.22	64.73	62.53	59.50
2.0σ	86.47	86.20	85.36	83.90	81.68	78.51
2.5σ	95.61	95.43	84.87	93.87	92.26	89.80
3.0σ	98.89	98.81	98.54	98.03	97.14	95.63
3.5σ	99.78	99.76	99.66	99.46	99.07	98.31
4.0σ	99.97	99.96	99.93	99.87	99.74	99.42

相関係数が大きくなるほど確率は下がる、分布の円が楕円に成るためで簡単にイメージできる。

相関係数が確率に影響する程度を認識出来れば充分と思います、細かい計算はプログラム（Book）がないと出来ませんので。

プログラム（Book）で計算するとメッシュの荒さの関係で1.5σから2σの当たりで0.01％位の計算誤差があるので注意ください。

相関係数と次の公差中心からのズレもそうですが基準点測量の影響をうけますので厳密網計算されること、網の振れが出来るだけでないように長い路線の単路線結合は避ける等の注意が必要のようです。

分布中心のズレと確率

公差中心から分布中心への距離

標準偏差	0	公差の1/10	公差の1/5	公差の1/3	公差の1/2.5	公差の1/2	1σ時の計算式との差
1.0σ	39.35%	39.23%	38.77%	37.66%	36.91%	36.03%	0.001%
1.5σ	67.53	67.26	66.26	63.48	62.23	57.63	0.010
2.0σ	86.46	85.98	84.25	81.32	78.71	73.85	0.008
2.5σ	95.60	95.12	94.00	90.46	88.39	82.95	0.002
3.0σ	98.89	98.70	98.06	95.82	93.88	90.11	0.000
3.5σ	99.78	99.70	99.38	98.09	96.99	93.89	0.000
4.0σ	99.97	99.94	99.85	99.30	98.59	96.83	0.000

分布の中心が公差の中心から離れればそれだけ確率は下がる。
標準偏差で3σ以上、公差からの距離が1／5以下でないと計算精度が出ないので注意する、0.01σメッシュ改良値。

公差の中心から分布の中心がずれていった場合、確率にどの程度の変化が出るのかを見てください。

公差10分の1とは公差が60mmの場合で6mmズレ場合です。

表右は0.01σメッシュで計算精度で計算式との差では3σ以上では全く問題は無い、1.5σ～2σ当たりで0.01％（一万分の一）ほど計算誤差があるがこの量域は合否の判定には使われない。

通常は4.25σ以上で使いますので全く問題はありません。

例題2

次のデータはある現地から20個の点をランダムに抽出したものである、公差を0.200としたときに全体のデータは 良 か 不良 か。

	図面値			実測値			
点番	点名	X	Y	点名	X	Y	位置誤差
1	C21	132.046	687.816	Z21	132.033	687.830	0.019
2	C1	71.177	707.981	Z1	71.168	708.004	0.025
3	C2	74.220	707.482	Z2	74.220	707.479	0.003
4	C3	79.379	701.531	Z3	79.396	701.535	0.017
5	C5	84.964	701.045	Z5	84.969	701.031	0.015
6	C6	93.857	701.111	Z6	93.863	701.099	0.013
7	C7	95.627	701.812	Z7	95.638	701.811	0.011
8							
9	C9	115.812	701.845	Z9	115.828	701.838	0.017
10	C19	135.027	705.821	Z19	135.183	705.776	0.162
11	C10	141.732	707.080	Z10	141.792	707.120	0.072
12	C11	144.517	706.140	Z11	144.581	706.188	0.080
13	C12	147.234	703.767	Z12	147.233	703.777	0.010
14							
15	C13	161.537	682.255	Z13	161.646	682.109	0.182
16							
17							
18	C22	137.608	666.511	Z22	137.638	666.515	0.030
19	K4	131.511	663.424	ZK4	131.510	663.429	0.005
20	K5	131.127	660.206	ZK5	131.144	660.202	0.017
21	K6	131.142	660.056	ZK6	131.161	660.052	0.019
22	K7	131.290	658.941	ZK7	131.289	658.947	0.006
23	K8	131.762	654.799	ZK8	131.768	654.798	0.006

では問題を解いて見てください。

普通は別表第5の甲2の基準でまず平均二乗誤差を計算します、次に各点の位置誤差が0.200以下かどうかで判断するわけです。

平均二乗誤差を計算すると64mmで平均二乗誤差公差の70mmに対しては「良」と判断されます。

回答2

答えは　不良　不良率0.15%(不良率の限度意を0.01%未満として)

二変量(位置)不良率計算表

限界値(10以下)	5.18	実行
標準偏差　σ	41	不良です
相関係数(0～<1)	0.52	元数値
位置誤差の公差	200	不良率設定%
公差からの距離　S	20	0.013
確　率　%	99.847	
異常率　%	0.153	10　←ズーム

重要なのは公差内であっても全体としては【不良】とされることです。

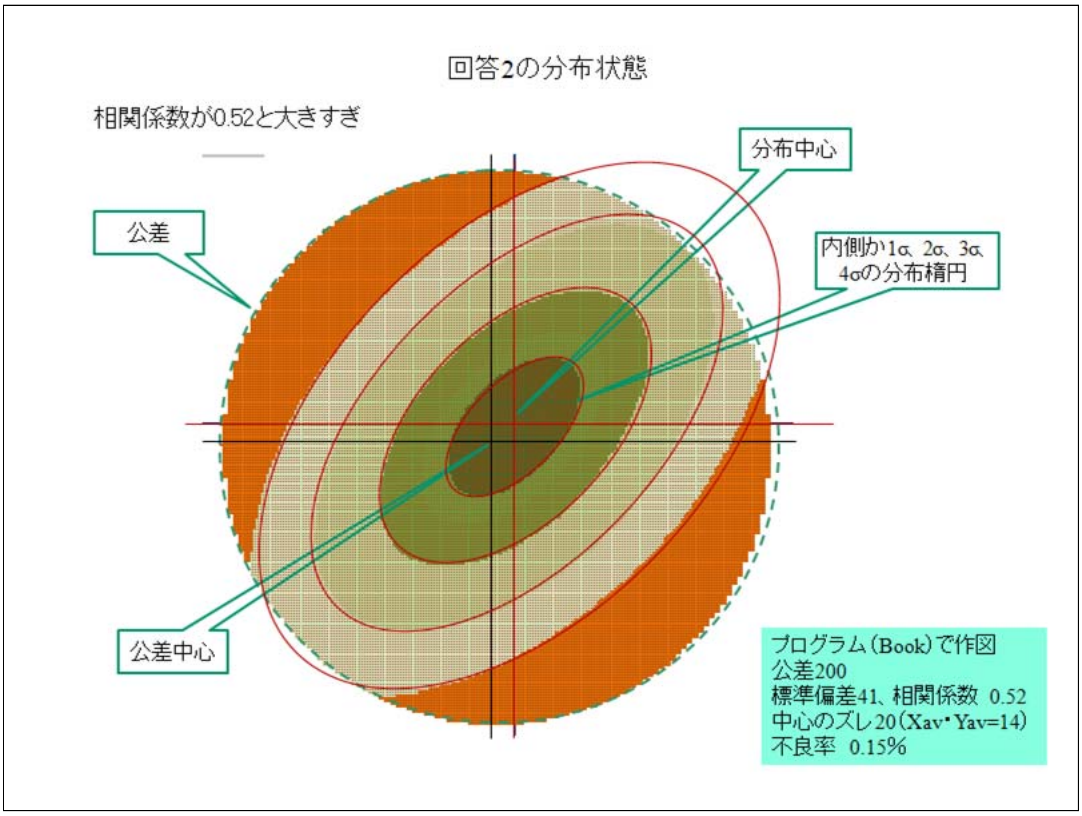
プログラム(Book)がないと計算出来ないでしょう
希望者は　http://012.o.oo7.jp/　からメールをくださいプログラム(エクセルBook)差し上げます。

平均二乗誤差は相関係数が0が前提になっており相関係数が大きなデータに対しては甘い基準になっています。

ここでは標準偏差、相関係数、公差からの距離S、公差の四要素から不良率を求めて判断します。

不良率の限度を0.01%とすれば0.15%で全体がNGと判断されます。

要するに分布中心のズレ、相関係数が大きく影響してると言うことです。



このデータでは相関係数（楕円の状態）が0.52と大きいことによって、分布の中心が公差の中心からズレていることから不良率が大きくなっている。

中心のズレは基準点がずれていることによって起こることがありますし網の組み方、計算方法によっても起きます。

そのことから言えばこの例では相関係数が大きい、通常は0.3以下なので0.52では大きすぎます。

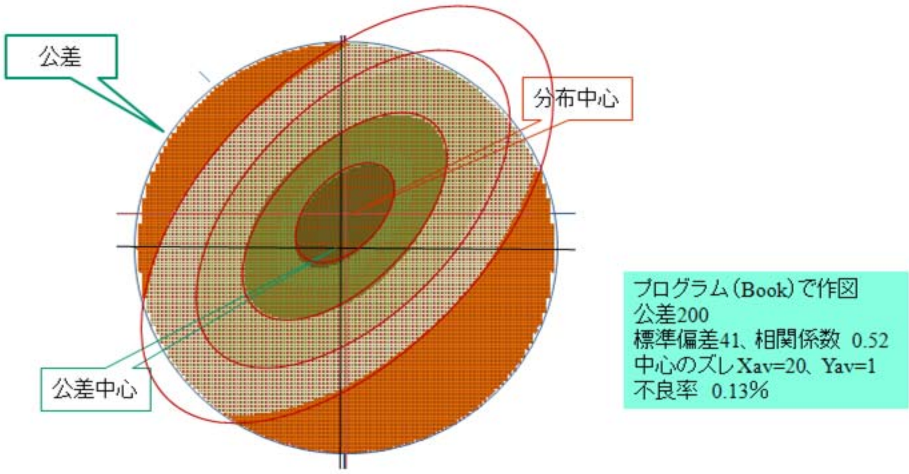
基準点測量で注意しなければならないのが同じ与点（3級程度）を使って4級とか登記基準点を展開する場合に簡易網平均で計算すると基準点が平行にずれた形で計算されることがある。

相関係数は古い図面ほど大きくなる傾向があることから測量精度の低いことと関係していると考えられる。

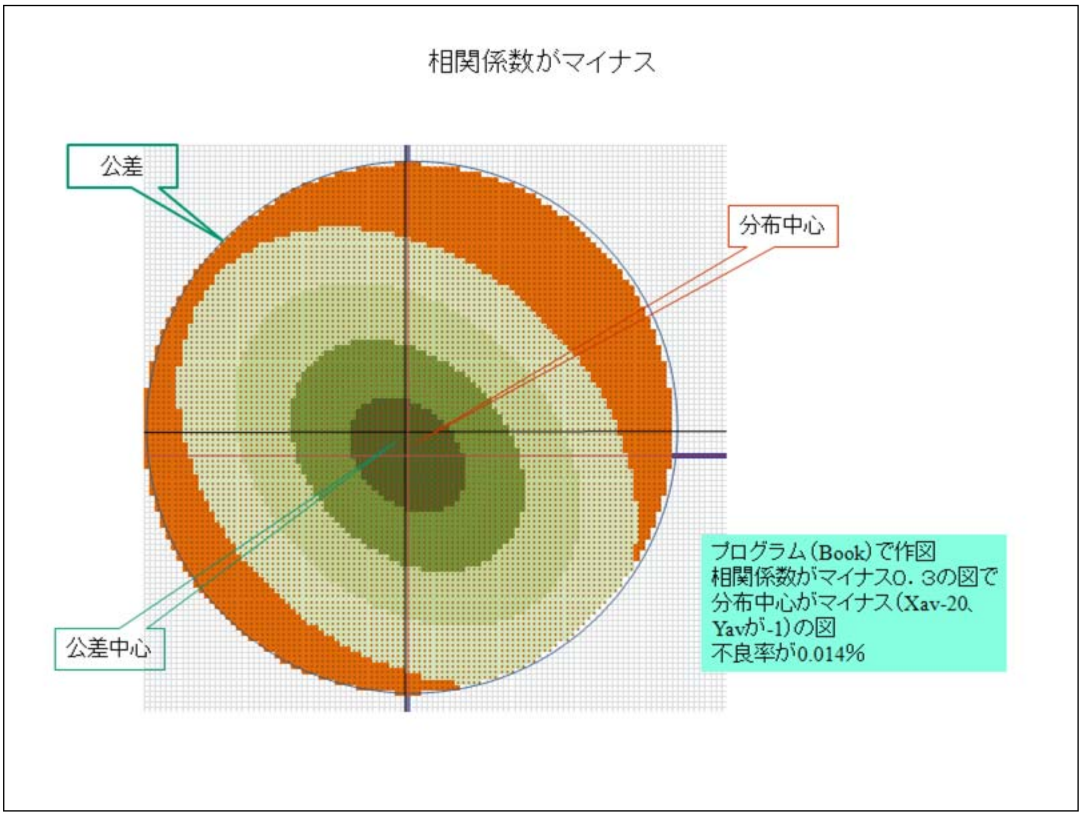
標準偏差、相関係数、中心のズレ、公差から不良率が計算できることを知って頂ければよいと思います。

分布中心のズレ

分布が45度の線上にある場合を仮定してきましたが実際の分布を考慮するとどうなるのか計算してみます。
場合によっては図の様な状態もあると言うことですが実際は極端な例は生じないでしょう、つまり公差中心の近くに分布中心があるのが普通です。

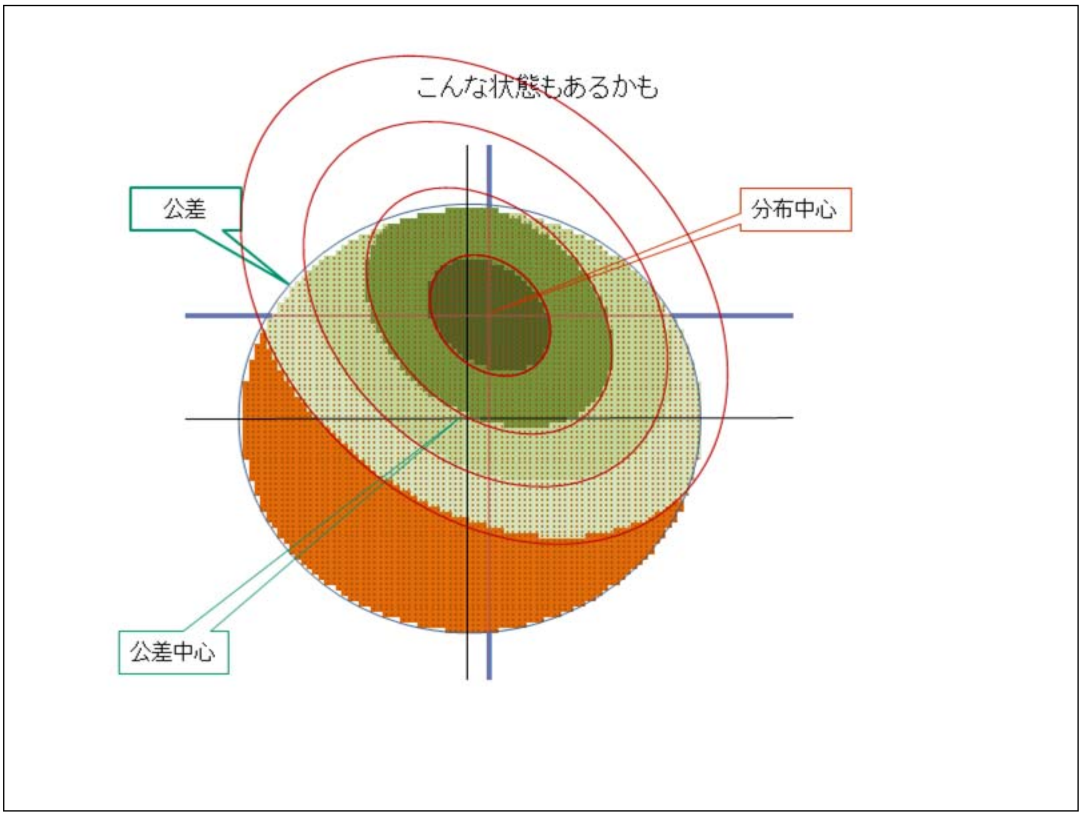


公差中心に対する分布中心の位置はXav（誤差の平均値）とYavで計算されますのでこの図の様になります、相関係数を求める関係で分布軸をプラナでは45度、マイナスでは135度に回転させているので分布中心のXavとYavも回転させなければならない。



ともかく分布のパターンはた多種多用で一定の法則は見あたらないので計算してみないと判らない。

今までは想像の域であった状態が計算でも分布図でもみられることが判った、不良率も計算できるこのことだけでも大きな進歩といえるのではないだろうか。

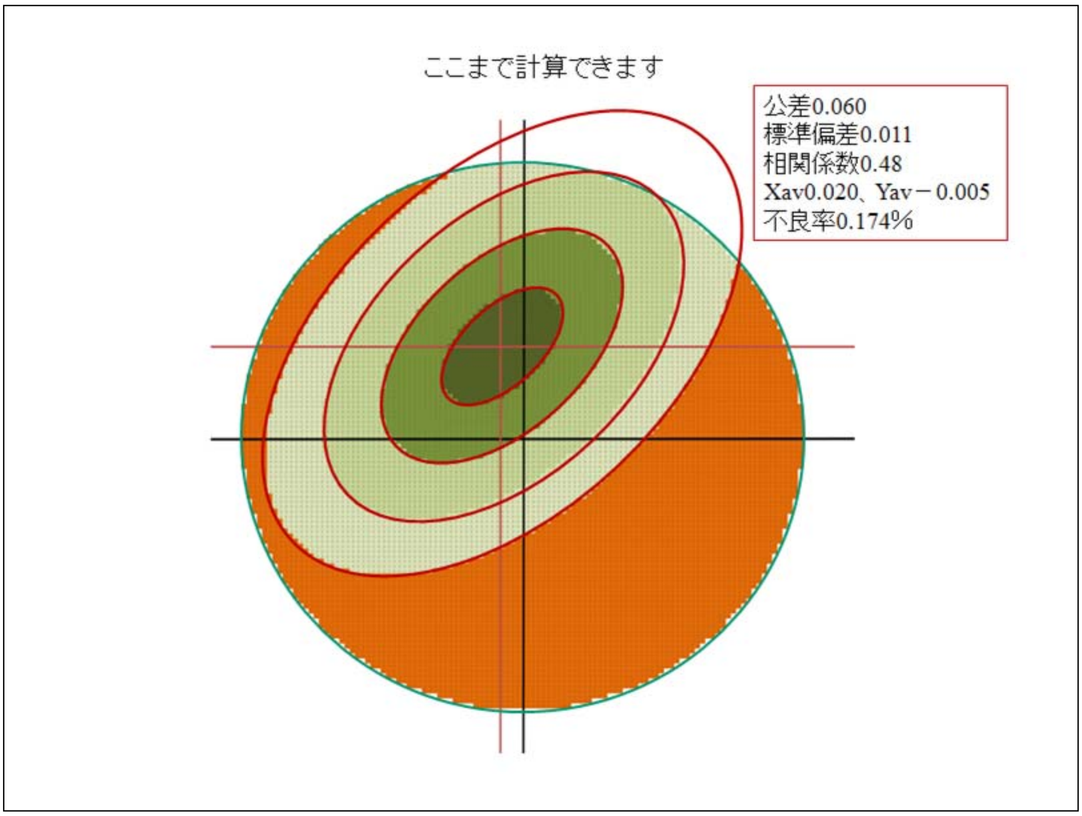


実際にはこんな極端な例はないでしょう・・・と思いますが計算してみれば判ります。

このように出来てしまうと様々なことが見えてくる、そうして以外と複雑なことも判ってきた。

そうして考えると平均二乗誤差を判定指数に考えた方は本当にすごいといえるが残念ながら正確ではない。

正確度という考え方もありますがそれはそれで判りにくいと言うが第一印象です。


ここまで計算できます
公差0.060
標準偏差0.011
相関係数0.48
Xav0.020、Yav－0.005
不良率0.174%