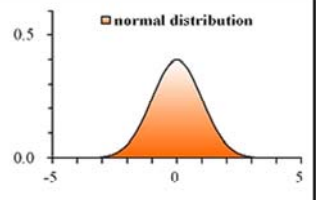
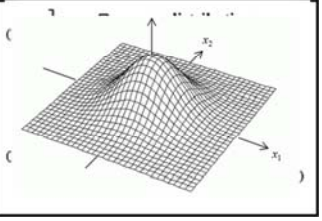
確率分布とは

確率変数(正規分布になるデータ)がある区間内の全ての実数(実データ)を取り得る場合は「連続型」といいます。連続型のグラフは，横軸の確率変数が連続量なので，縦軸はその値での確率密度を表しており，区間内(横軸のある値とある値の間)を積分した面積，二変量分布では体積がその確率に相当します。一変量では釣鐘型分布の内側を100としたに指定した範囲の面積を確率と言う，二変量では範囲の体積を確率という。

連続型確率分布

境界(筆界)復元では下表の正規分布，上側の一変量の分布と下の多変量の分布，境界(筆界)復元では多変量のうち二変量の分布を使います。この二つの分布を理解すれば境界(筆界)復元に出てくる様々な状況に対応出来ます。

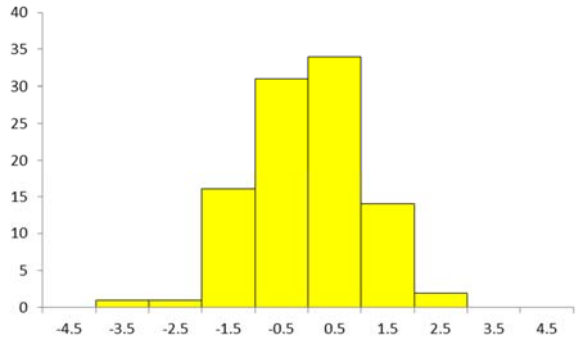
確率分布の名称	母数(パラメータ)	確率変数Xとその範囲(区間)	単峰形		単調形		その他の分布形	分布形(グラフ)の一例
			対称	歪み	減少	増加		
正規分布 (ガウス分布) $N(\mu, \sigma^2)$	平均 μ， 分散 σ^2	X：実数， $-\infty < X < \infty$						
多変量正規分布	期待値ベクトル， 分散共分散行列	Xi：実数 $-\infty < Xi < \infty$	多次元型 (正規分布の多変量化)					

正規分布

データをいくつかの階級に分けて度数分布表やヒストグラムを作成したとき，中心付近の度数が最も高くなり，そこから左右に同程度で度数が少なくなっていく形になります，測定誤差や自然現象の中に現れるバラツキは正規分布に従うと見なせるものが多く，統計学の理論上も境界(筆界)復元などの応用上も非常に重要で実用性の高い分布です。

境界(筆界)復元測量は長年に渡ってその手法が確立されてきた分野であり面積の精度，辺長の精度，境界点位置の精度から正規分布が見られる確率分布です。19 世紀に伊能忠敬が日本列島の地図作成に全国を測量していたほぼ同時代にドイツのガウスが観測誤差の研究から導いたことが有名であることからガウス分布と呼ばれることもあります。

下図は横軸に誤差の大きさ，縦軸にその量をグラフにしたもので，この説明を理解できると思います。



正規分布の形，正規分布は $N(\mu, \sigma^2)$ と表記します，これはカッコ内の2つの値， 平均 μ と 分散 σ^2 が決まれば正規分布が一意に定まることを意味しており，この平均 μ と 分散 σ^2 を母数といいます。

正規分布は平均 μ を中心として左右対称になった西洋の釣鐘と似た形状の曲線（ベルカーブ，釣鐘型）の分布形状を描きます。

平均 μ は分布形の中心的位置を表しているので，平均の違いは位置の違いとして表れます，また，分散 σ^2 （標準偏差 σ ）については，その値が大きくなるほど釣鐘型の曲線が横に伸びて裾野が広がる形になりますが，これは形が横に伸びただけで，正規分布の曲線の本質的な形状は，相対的に一定で決まった形をしています。

境界（筆界）復元では比較ずる境界図が同一の座標軸，測量器機，測量方法等によって作成されていないため座標変換をした上で比較しますので平均 $\mu=0$ として度数分布表やヒストグラムを作成します。

同一与点（境界測量の基準とする点）で点検する場合，地籍調査，地図作成などの場合は座標変換をせずに比較判断しますので平均 $\mu=0$ にはなりません。

連続型の確率変数 X が正規分布 $N(\mu, \sigma^2)$ に従うとき，その確率密度関数 $f(x)$ は

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

で表されます，確率変数 X は $-\infty < x < +\infty$ の範囲の実数をとります．この $f(x)$ は $x = \mu$ のときが最大値であり， $x = \mu \pm \sigma$ の点が変曲点となっています。

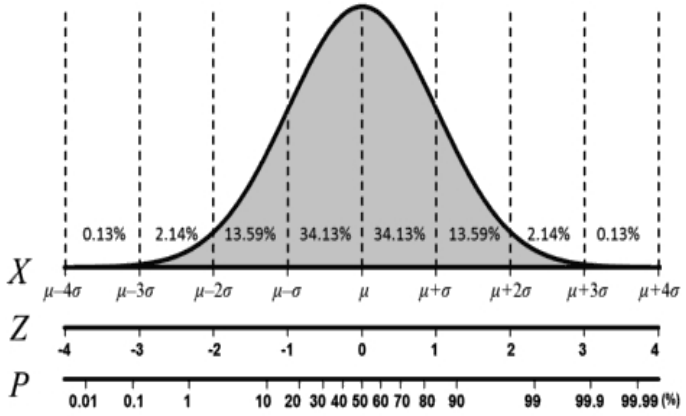
正規分布に限らず，どのような分布であっても，平均 μ は確率変数 X の分布における位置を示す指標であり，標準偏差 σ は分布における確率変数 X のバラつきの尺度となります。

さまざまな分布が持っているこの平均と標準偏差の違いを，何らかの標準的な形に変換することができれば，さまざまな分布の姿を一定の基準で比較・検証することも可能となります。

そこで，位置の基準である平均を 0 ，尺度である標準偏差を 1 に変換することを考えます。確率変数 X を次式で変換すると，変換された確率変数 Z は，平均が 0 ，標準偏差が 1 の分布になります。

基準となる標準正規分布，平均 μ ，標準偏差 σ の正規分布 $N(\mu, \sigma^2)$ に従う確率変数 X を標準化変換した確率変数 Z は，平均が 0 ，標準偏差が 1 の正規分布 N (0, 1) に従うことになります。この N (0, 1) を特に標準正規分布といいます。

標準正規分布 N (0, 1) は， $z = 0$ の点を中心（平均）とした形です．したがって， $z = 0$ で確率は半々ですから $P = 50\%$ となります．確率変数 X のある値 x を標準化変換した z の意味は，元の一般的な正規分布の値 x が，平均 μ から標準偏差 σ の z 倍だけ離れているということを示しています。 標準正規分布 N (0, 1) に従う確率変数 Z の確率密度関数を ，累積分布関数（下側確率）を と表すと次図のようになります。



確率密度関数について，負の無限大から z までの範囲を積分計算したもの（面積）が累積分布関数（下側確率）です（図を参照）。

累積分布関数の積分は Excel 等のプログラムを用いて求めることになります。計算されたものを数表としてまとめた 正規分布表 が次表です，また主な表計算ソフト（Excel）には標準正規分布に関する関数が用意されています。

次の表は Excel で作成したものです，左から標準偏差の倍数，二変量（2 次元）の確率，一変量（1次元）の両側確率，一変量（1次元）の片側確率を示します。

3行目の標準偏差に標準偏差の倍率を入力すればその標準偏差の確率が計算できるようにプログラムされています。

以前であれば統計学の書籍の巻末に表が添付されていてその表から確率を求めていましたがいまでは様々なプログラムによって簡単に求まられます。

	2変量	1変量	1変量
標準偏差	確率	両側確率	片側確率
2.576	0.963771	0.990005	0.495002
0.05	0.001249	0.039878	0.019939
0.1	0.004988	0.079656	0.039828
0.15	0.011187	0.119235	0.059618
0.2	0.019801	0.158519	0.079260
0.25	0.030767	0.197413	0.098706
0.3	0.044003	0.235823	0.117911
0.35	0.059412	0.273661	0.136831
0.4	0.076884	0.310843	0.155422
0.45	0.096293	0.347290	0.173645
0.5	0.117503	0.382925	0.191462
0.55	0.140367	0.417681	0.208840
0.6	0.164730	0.451494	0.225747
0.65	0.190428	0.484308	0.242154
0.7	0.217295	0.516073	0.258036
0.75	0.245160	0.546745	0.273373
0.8	0.273851	0.576289	0.288145
0.85	0.303195	0.604675	0.302337
0.9	0.333023	0.631880	0.315940
0.95	0.363168	0.657888	0.328944
1	0.393469	0.682689	0.341345
1.05	0.423771	0.706282	0.353141
1.1	0.453926	0.728668	0.364334
1.15	0.483794	0.749856	0.374928
1.2	0.513248	0.769861	0.384930
1.25	0.542167	0.788700	0.394350
1.3	0.570443	0.806399	0.403200
1.35	0.597979	0.822984	0.411492
1.4	0.624689	0.838487	0.419243
1.45	0.650499	0.852941	0.426471
1.5	0.675348	0.866386	0.433193
1.55	0.699182	0.878858	0.439429
1.6	0.721963	0.890401	0.445201
1.65	0.743660	0.901057	0.450529
1.7	0.764254	0.910869	0.455435
1.75	0.783735	0.919882	0.459941
1.8	0.802101	0.928139	0.464070
1.85	0.819360	0.935686	0.467843
1.9	0.835526	0.942567	0.471283
1.95	0.850618	0.948824	0.474412
2	0.864665	0.954500	0.477250
2.05	0.877697	0.959636	0.479818
2.1	0.889749	0.964271	0.482136
2.15	0.900863	0.968445	0.484222
2.2	0.911078	0.972193	0.486097
2.25	0.920440	0.975551	0.487776
2.3	0.928995	0.978552	0.489276
2.35	0.936787	0.981227	0.490613
2.4	0.943865	0.983605	0.491802
2.45	0.950275	0.985714	0.492857
2.5	0.956063	0.987581	0.493790
2.55	0.961274	0.989228	0.494614
2.6	0.965953	0.990678	0.495339
2.65	0.970140	0.991951	0.495975
2.7	0.973879	0.993066	0.496533
2.75	0.977206	0.994040	0.497020
2.8	0.980159	0.994890	0.497445
2.85	0.982773	0.995628	0.497814
2.9	0.985079	0.996268	0.498134
2.95	0.987109	0.996822	0.498411
3	0.988891	0.997300	0.498650
3.05	0.990450	0.997712	0.498856
3.1	0.991811	0.998065	0.499032
3.15	0.992996	0.998367	0.499184
3.2	0.994024	0.998626	0.499313
3.25	0.994914	0.998846	0.499423
3.3	0.995682	0.999033	0.499517
3.35	0.996344	0.999192	0.499596
3.4	0.996911	0.999326	0.499663
3.45	0.997397	0.999439	0.499720
3.5	0.997813	0.999535	0.499767
3.55	0.998166	0.999615	0.499807
3.6	0.998466	0.999682	0.499841
3.65	0.998720	0.999738	0.499869
3.7	0.998935	0.999784	0.499892
3.75	0.999116	0.999823	0.499912
3.8	0.999268	0.999855	0.499928
3.85	0.999396	0.999882	0.499941
3.9	0.999502	0.999904	0.499952
3.95	0.999591	0.999922	0.499961
4	0.999665	0.999937	0.499968
4.05	0.999726	0.999949	0.499974
4.1	0.999776	0.999959	0.499979
4.15	0.999818	0.999967	0.499983
4.2	0.999852	0.999973	0.499987
4.25	0.999880	0.999979	0.499989
4.3	0.999903	0.999983	0.499991
4.35	0.999922	0.999986	0.499993
4.4	0.999937	0.999989	0.499995
4.45	0.999950	0.999991	0.499996
4.5	0.999960	0.999993	0.499997
4.55	0.999968	0.999995	0.499997
4.6	0.999975	0.999996	0.499998
4.65	0.999980	0.999997	0.499998
4.7	0.999984	0.999997	0.499999
4.75	0.999987	0.999998	0.499999
4.8	0.999990	0.999998	0.499999
4.85	0.999992	0.999999	0.499999
4.9	0.999994	0.999999	0.500000

この表の要点を示せば下表のとおりです，下表の上は一般的に説明される標準偏差と確率の関係です。

下表の下は有意水準に使われる標準偏差の値を示しました，境界（筆界）復元計算では $\mu \pm 1.96\sigma$，信頼率 95%，有意水準 5%を使います。

一変量

区間の幅	μ ± 1σ	μ ± 2σ	μ ± 3σ	μ ±3.89σ
区間に入る確率	68.27%	95.45%	99.73%	99.99%
区間から外れる確率	31.74%	4.56%	0.26%	0.01%
区間から外れる割合	約1/3	約1/20	約1/400	約1/10000

区間の幅	μ ± 1.96σ	μ ± 2.58σ	μ ± 2.81σ	μ ±3.29σ
区間に入る確率	95.00%	99.00%	95.50%	99.90%
区間から外れる確率	5.00%	1.00%	0.50%	0.10%

ヒストグラムの作成

正規分布からデータの全体像がわかるためにはヒストグラムを作成することにより理解が出来ます。データを分析する上でとても有用な手段であります，それはどのようなデータでも分布の形が対称で単峰と見なせる場合には，「平均を中心」に「標準偏差を尺度」として見ることにより，大まかですがデータの全体像がわかるということです。

境界（筆界）復元をする場合は標準偏差から誤差の全体像をつかめるようになっていることが求まられます，ただ単純に数値が小さければ良いというものではありません。

まず，はじめに誤差の分布を理解する上で必要なヒストグラムの作成について説明します，ヒストグラムの作成方法は様々ありますがここではガウスの正規分布曲線を意識した方法で行います。実データにおいて，平均を中心に標準偏差　σ　がプラスマイナス何個分の区間だと何パーセントの割合であるのかをヒストグラムから数値をイメージできれば，標準偏差から全体像をつかみやすくなります。

ヒストグラムを作ること自体が標準偏差，正規分布，確率を理解する上で役にたちますので作って見ることを勧めます。

はじめに下表の一連の観測データから標準偏差を計算します。

観測値
15.0100
15.1200
15.0201
15.0800
15.1100
15.0900
15.0501
15.0700
15.0502
15.0200
15.0503
15.0400
15.0301
15.0300
15.1500

標準偏差の計算は観測データから直接計算する方法と平均値との差，較差から計算する方法があります，当然結果は同じです。

実際にはエクセルの関数を使って計算します，関数は STDEV(A1:A15) で( )内はデータのあるセルの範囲です。また関数は STDEVP(A1:A15) でもかまいません，使い分けはデータのバラツキを知りたい場合は STDEV(A1:A15) で 計算式の中で計算重量としての使う場合は STDEVP(A1:A15) といった理解で良いと思います。

標準偏差の計算式は次のとおりで，記号符号の使い方は書籍によって違いがありますが式は同じです，実際には EXCEL 等で計算しますのでこの部分を丁寧に説明する意味があるかどうかは判りませんが，説明します。

観測値： X_i

平均値:$\overline{X}$

較差:$\Delta X_i = X_i - \overline{X}$

データ数： n

合計： $[\Delta X_i^2]$ i＝1～n 個の合計

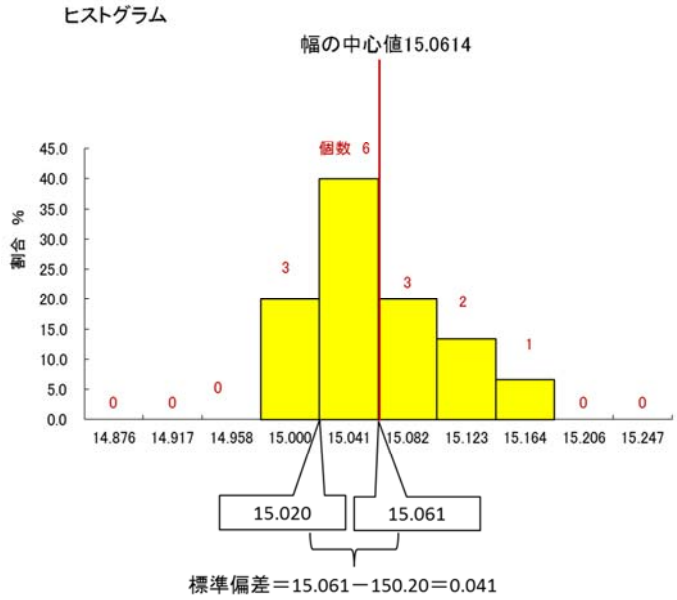
標準偏差:$\sigma = \sqrt{\frac{[\Delta X_i^2]}{n-1}}$

このときの分母の $n-1$ が STDEV(A1:A15) で 分母が n が STDEVP(A1:A15)になります。

観測データから直接計算する方法

	観測値	平均値	観測値－平均値	
	Xi	$\overline{X}$	$\Delta X = Xi - \overline{X}$	ΔX^2
1	15.0100	15.0614	−0.0514	0.0026406
2	15.1200		0.0586	0.0034355
3	15.0201		−0.0413	0.0017046
4	15.0800		0.0186	0.0003465
5	15.1100		0.0486	0.0023633
6	15.0900		0.0286	0.0008187
7	15.0501		−0.0113	0.0001274
8	15.0700		0.0086	0.0000742
9	15.0502		−0.0112	0.0001251
10	15.0200		−0.0414	0.0017129
11	15.0503		−0.0111	0.0001229
12	15.0400		−0.0214	0.0004574
13	15.0301		−0.0313	0.0009789
14	15.0300		−0.0314	0.0009851
15	15.1500		0.0886	0.0078523
			$[(\Delta X^2)]$ =	0.0237453

標準偏差:$\sigma = \sqrt{\frac{[\Delta X_i^2]}{n-1}} = \sqrt{\frac{0.0237453}{15-1}} = 0.04118$



同じデータを較差(観測値－平均値)で計算しますと次の表になります，正規分布に限らず，どのような分布であっても，平均， は確率変数 X の分布における位置を示す指標であり，標準偏差 σ は分布における確率変数 X のバラつきの尺度となります。 さまざまな分布が持っているこの平均と標準偏差の違いを，何らかの標準的な形に変換することができれば，さまざまな分布の姿を一定の基準で比較・検証することも可能となります。

そこで，位置の基準である平均を 0 ，尺度である標準偏差を 1 に変換することを考えます。確率変数 X を下表で変換すると，変換された確率変数は，平均が 0 ，標準偏差が 1 の分布になります。

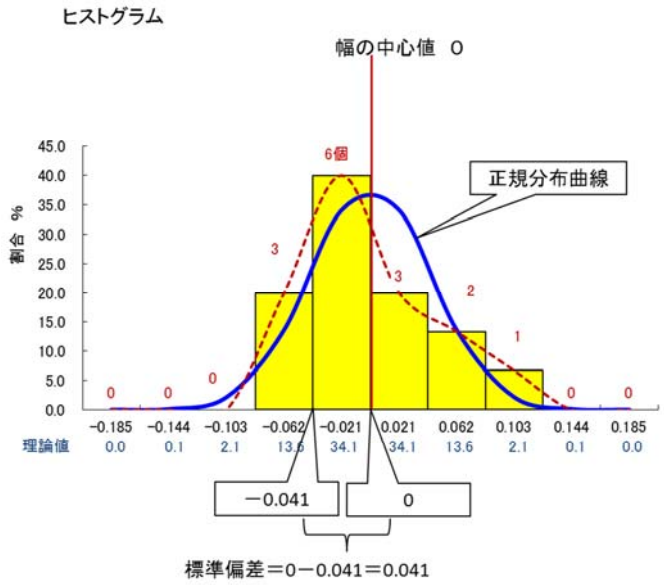
較差から計算する方法

	観測値 Xi	平均値 $\bar{X}$	較差 $\Delta X = Xi - \bar{X}$	較差の平均値 $\bar{\Delta X}$	$\Delta X - \bar{\Delta X}$	ΔX^2
1	15.0100	15.0614	−0.0514	0.0000	−0.0514	0.0026406
2	15.1200		0.0586		0.0586	0.0034355
3	15.0201		−0.0413		−0.0413	0.0017046
4	15.0800		0.0186		0.0186	0.0003465
5	15.1100		0.0486		0.0486	0.0023633
6	15.0900		0.0286		0.0286	0.0008187
7	15.0501		−0.0113		−0.0113	0.0001274
8	15.0700		0.0086		0.0086	0.0000742
9	15.0502		−0.0112		−0.0112	0.0001251
10	15.0200		−0.0414		−0.0414	0.0017129
11	15.0503		−0.0111		−0.0111	0.0001229
12	15.0400		−0.0214		−0.0214	0.0004574
13	15.0301		−0.0313		−0.0313	0.0009789
14	15.0300		−0.0314		−0.0314	0.0009851
15	15.1500		0.0886		0.0886	0.0078523
					$[(\Delta X^2)]$ =	0.0237453

次のグラフは幅の中心値(誤差の平均値)を 0 として，そこから左右側に標準偏差の幅だけの枠を設け，その枠の中に上の表にある⊿X が何個有るかをカウントし，その割合を%にした

棒グラフです，実線の曲線は正規分布の時の割合を棒グラフでなく，曲線で表示したもので正規分布曲線といいます。

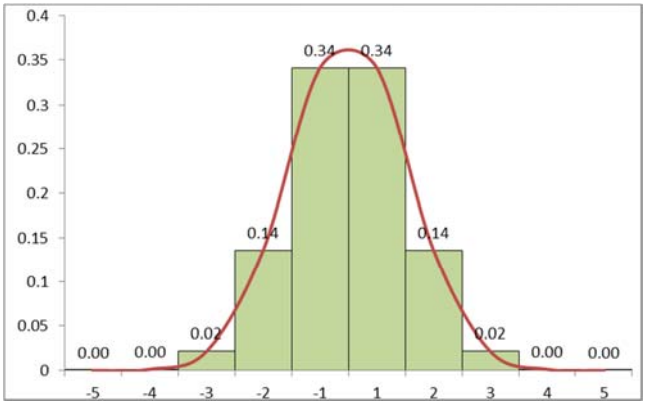
点線の曲線はカウントした割合を棒グラフでなく曲線で結んだもので，実際の分布曲線になります。こうすれば正規分布曲線と実際の分布曲線のズレ，不一致の状態が理解しやすいでしょう。この曲線がどのような形で一致してくれればよいのかは「検定」のところで説明します。



実際のデータ数には限りがあります，その結果理論上の正規分布曲線に実際の分布曲線が一致しないもので，その説明を次にします。

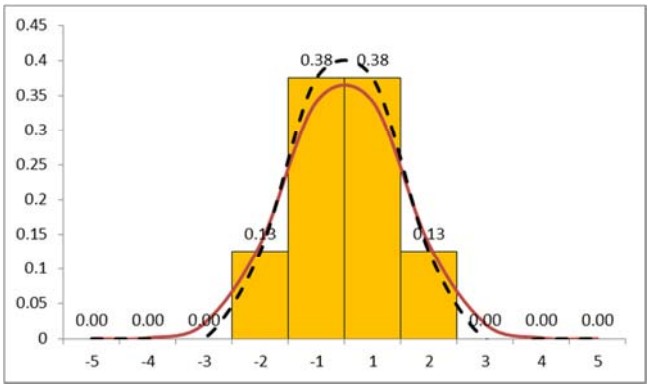
次のグラフは理論上の確率をヒストグラムと正規分布曲線で書いたものです，棒グラフの頂点中心を曲線が通っていいます。グラフの幅を1倍標準偏差としてあります。データ数によって正規分布曲線と実分布曲線を一致させる事は出来ません，課題となるのは分布曲線の裾の部分の不一致です。

裾の部分で正規分布曲線に対して実分布曲線が小さい値であれば問題にしません，問題なのはその逆の場合，正規分布曲線に対して実曲線が膨らんでいる状態です。



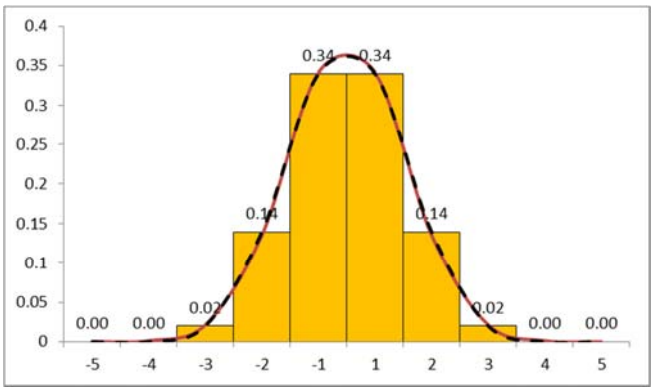
次のグラフはデータ数を10個にしたときに出来る正規分布データのヒストグラムです，曲線

は正規分布曲線，点線は10個のデータ曲線で，これ以上に最適な組合せはあり得ない状態ですが正規分布曲線とは山の頂点付近と裾付近で多少ズレがあります。データ数を偶数にして左右対称になるようにしてあります。



次にデータ数を100個にしたときのグラフを見てみます。最初の正規分布のグラフと一致しています，これはデータ数が相当数なければならないということです。

実際の境界図ではこのデータ数を確保することが図面上あるいはコスト上困難なことが多くありますので少ないデータで正規分布に近いか，遠いかの判断が難しくなります。



二変量の確率分布

境界（筆界）復元においてはＸ座標値，Ｙ座標値から較差（⊿x ・⊿y） を求めて散布図を作成すれば位置の基準である平均を ⊿x=0 ，⊿y＝0 の尺度である標準偏差を 1 に変換することによって確率変数，平均が 0 ，標準偏差が 1 の分布になります。

次表はその場合の座標値です，このように座標軸が異なることが普通なので座標変換によって回転，移動，伸縮によって重ねた座標値から⊿x ・⊿y を求めて散布図を作成します。

⊿x ・⊿y の値は省略してあります。このデータは観測の生データで分布上，異常な値，つまり正規分布に馴染まない値を除く等の処理をしていません。そのため以下に示す分布図の形状が正規分布から離れていることを承知の上ご覧下さい。

順	点名	X	Y	順	点名	X	Y
1	721	-54850.090	-25272.250	1	8	502.213	498.409
2	722	-54819.390	-25337.110	2	110	536.890	435.547
3	723	-54819.040	-25337.720	3	109	537.276	434.952
4	724	-54818.410	-25338.880	4	108	537.981	433.843
5	725	-54810.570	-25352.680	5	107	546.646	420.557
6	726	-54810.330	-25353.170	6	106	546.918	420.093
7	727	-54796.350	-25380.810	7	99	562.607	393.365
8	728	-54791.140	-25391.100	8	98	568.436	383.435
9	729	-54791.650	-25392.110	9	97	568.001	382.388
10	793	-54786.160	-25403.090	10	96	574.178	371.768
11	650	-54766.970	-25393.380	11	88	592.726	382.662
12	649	-54765.050	-25398.260	12	118	594.936	377.906
13	647	-54763.260	-25402.640	13	119	596.985	373.617
14	648	-54748.550	-25399.250	14	117	611.446	377.935
15	639	-54739.590	-25417.500	15	116	621.543	360.274
16	638	-54741.670	-25418.270	16	85	619.525	359.364
17	637	-54734.600	-25432.560	17	163	627.468	345.547
18	634	-54731.120	-25439.470	18	86	631.362	338.884
19	633	-54718.530	-25435.060	19	84	643.645	344.072
20	632	-54722.820	-25420.320	20	83	638.459	358.516
21	631	-54708.820	-25416.680	21	82	652.199	363.008
22	630	-54687.070	-25411.040	22	80	673.570	369.991
23	629	-54686.670	-25413.730	23	81	674.132	367.333
24	619	-54677.690	-25411.510	24	166	682.960	370.096
25	618	-54646.540	-25403.910	25	314	713.550	379.669
26	617	-54645.790	-25406.160	26	313	714.439	377.471
27	616	-54644.520	-25416.330	27	312	716.343	367.400
28	615	-54644.150	-25419.060	28	311	716.883	364.698
29	608	-54640.240	-25421.180	29	164	720.902	362.792
30	607	-54631.810	-25416.900	30	70	728.978	367.706
31	606	-54623.270	-25413.170	31	68	737.259	371.993
32	605	-54620.870	-25410.490	32	65	739.504	374.835
33	604	-54615.230	-25401.080	33	63	744.497	384.588
34	603	-54609.610	-25391.230	34	61	749.530	394.762
35	602	-54599.910	-25380.850	35	157	758.662	405.645
36	416	-54590.140	-25371.890	36	60	767.931	415.106
37	415	-54591.960	-25366.500	37	153	765.738	420.379
38	414	-54597.240	-25347.120	38	58	759.253	439.410
39	413	-54601.830	-25330.980	39	151	753.503	455.174
40	412	-54601.060	-25330.720	40	150	754.269	455.476
41	411	-54607.900	-25307.570	41	53	746.170	478.176
42	410	-54613.510	-25288.510	42	148	739.284	496.853
43	409	-54619.390	-25268.800	43	147	732.212	516.167
44	408	-54624.940	-25250.020	44	144	725.527	534.574
45	755	-54631.590	-25230.830	45	143	717.733	553.329
46	819	-54656.880	-25219.730	46	133	691.827	562.848
47	784	-54659.790	-25220.060	47	32	688.934	562.339
48	672	-54677.700	-25223.660	48	30	671.290	557.613
49	671	-54692.610	-25227.290	49	26	656.640	553.061
50	670	-54700.780	-25229.320	50	25	648.619	550.527
51	669	-54712.960	-25232.750	51	24	636.661	546.351
52	668	-54730.460	-25233.980	52	22	619.265	544.053
53	667	-54735.870	-25233.680	53	19	613.843	544.038
54	666	-54738.640	-25233.690	54	18	611.088	543.815
55	665	-54785.750	-25254.970	55	16	565.393	519.656
56	664	-54793.400	-25255.860	56	13	557.806	518.304
57	663	-54804.370	-25259.250	57	11	547.064	514.235

境界図から得られる二変量の分布は楕円になります，その結果，得られる指標は下表のとおりで，誤差楕円の長軸標準偏差σｍ・短軸標準偏差σｎ，相関係数（ρ），分布楕円の角度，二変量の標準偏差σです，境界(筆界)復元では楕円角度については重要ではありませんが相関係数を計算する上で必要です。

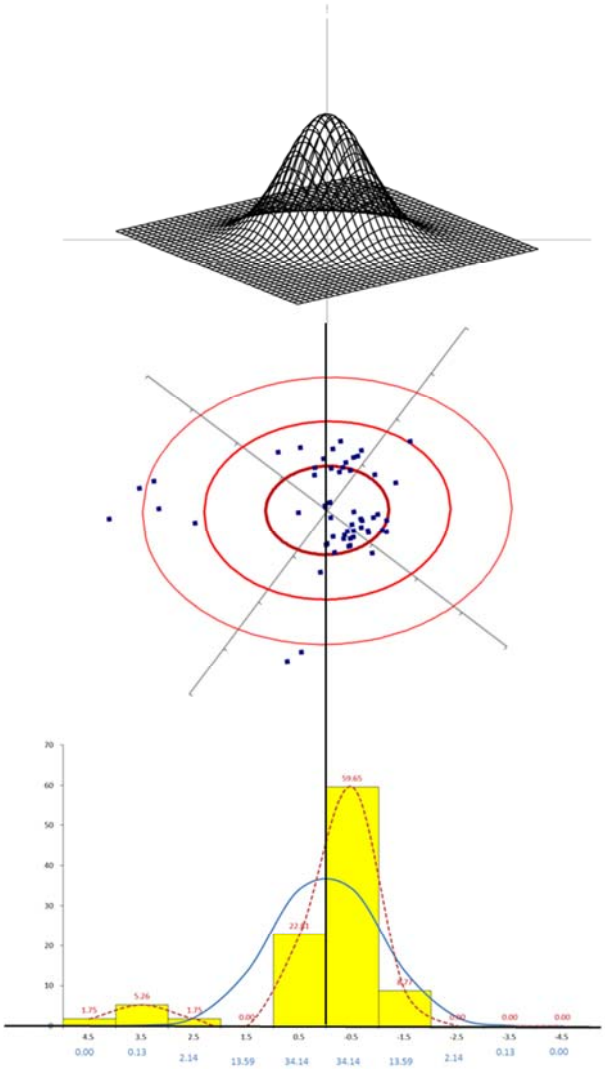
分布データ

σ_m	σ_n	$\overline{\Delta x}$	$\overline{\Delta y}$	相関係数
0.0543	0.0390	0.0000	0.0000	−0.320
σｘ	σｙ	σ	カウント	楕円角度
0.0493	0.0452	0.0473	57	−37

下図はその概要で，上から二変量の密度関数グラフ，この密度間数グラフはあらゆる角度から観測出来るようにプログラムしてあります。グラフを上 45° から俯瞰し，0 から 180 度まで 30 度毎に表しました。

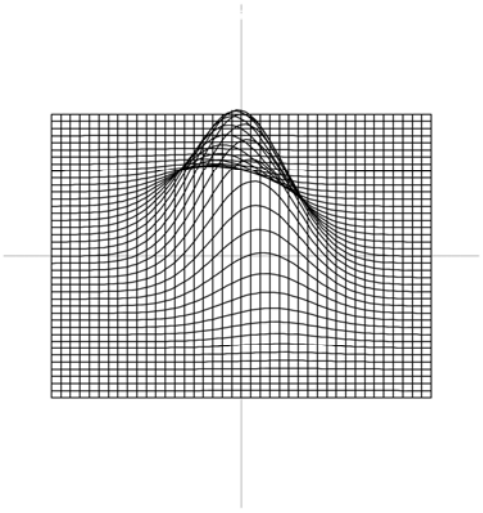
一変量は平面の釣鐘型ですが二変量は立体の釣鐘型になります，その下は散布図で分布の状態が楕円で計算できます，内側から 1 倍標準偏差，2倍標準偏差，3倍標準偏差となっています。

密度関数グラフの下の図は散布図を下から上に向かって見上げた時の座標軸毎のヒストグラムです，このヒストグラムは散布図の回転量によって変化します。0 から 180 度まで 30 度毎のヒストグラムをその下に表らします。

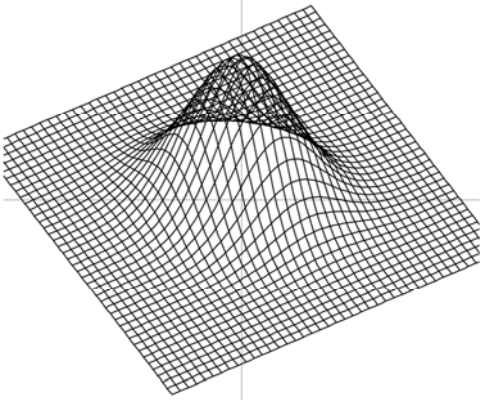


上から，密度関数グラフ，散布図，ヒストグラム（青の曲線は正規分布曲線，点線は実際の分布曲線）。

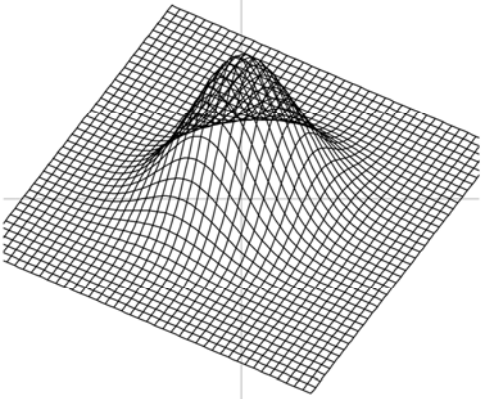
座標軸毎の密度関数グラフ



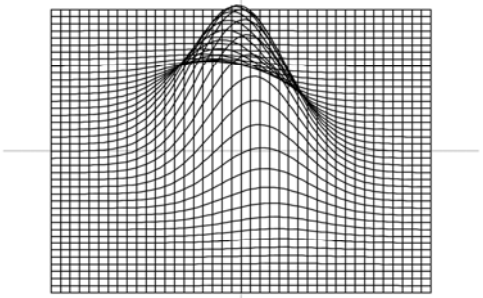
0 度



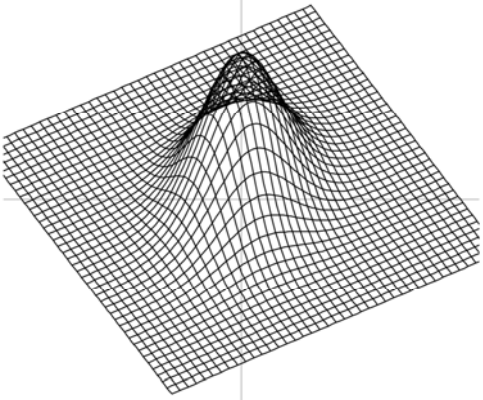
30 度



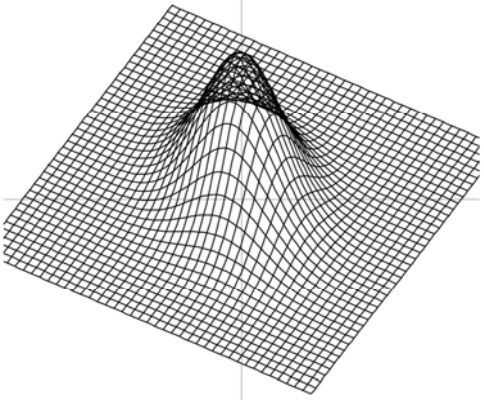
60 度



90 度

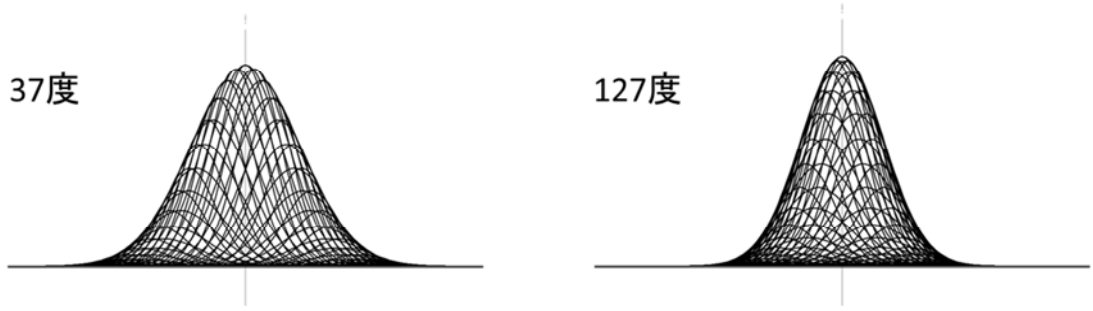


120 度

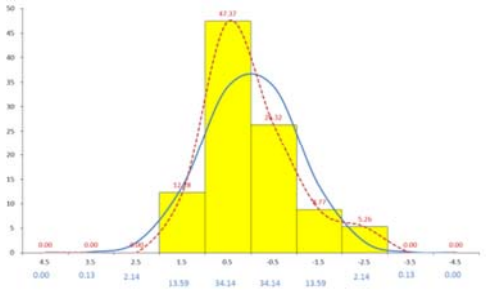


150 度

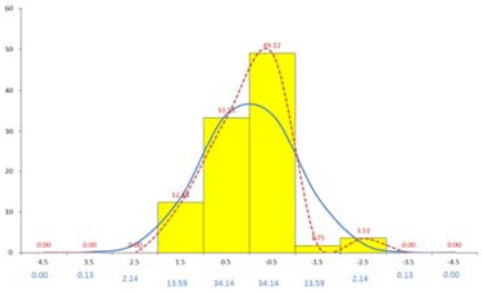
横方向，37 度（短軸角）と 127 度（長軸角）から見たグラフを参考までに示します。分布上の異常点を除いたからといってもこの形状が円形に成ると言うことではありません。



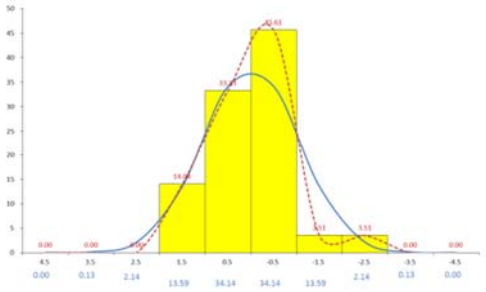
座標軸毎のヒストグラム



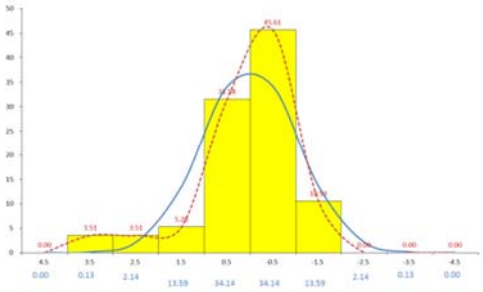
0 度



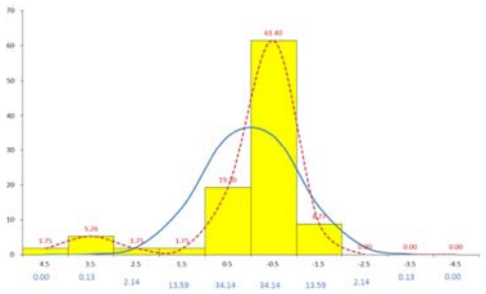
30 度



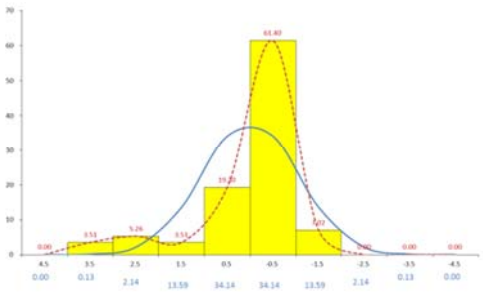
60 度



90 度



120 度



150 度

確率は立体の体積になりますので下表のように一変量とは異なります。

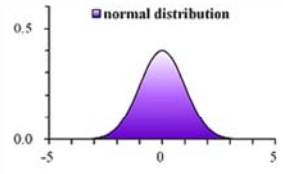
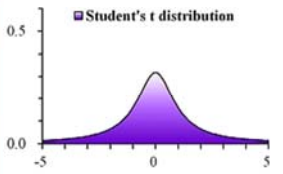
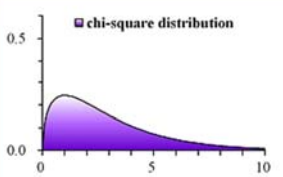
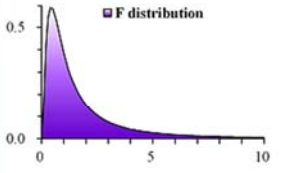
二変量

区間の幅	μ ± 1σ	μ ± 2σ	μ ± 3σ	μ ±4.30σ
区間に入る確率	39.35%	86.47%	98.89%	99.99%
区間から外れる確率	60.65%	13.53%	1.02%	0.01%
区間から外れる割合	約2/3	約1/6	約1/100	約1/10000

母数推定や仮説検定などの推測統計で用いられる確率分布

何を推定・検定するのかによって，それぞれの目的に応じた確率分布があります。推定・検定には多くの手法がありますが，ここでは，境界（筆界）復元に使われる推定や検定でよく用いられる代表的な確率分布とその統計量の一例を次に示します。

推定については「標準偏差の信頼区間・平均値の信頼区間」を参照して下さい，検定については「t異常値検定」「χ^2二乗正規分布適合度検定」「混合分布・層別・F検定」を参照して下さい。

確率分布の名称	母数（パラメータ）	確率変数Xとその範囲（区間）	推定・検定		分布形（グラフ）の一例
			統計量	用途の一例	
正規分布 N（0，1）	平均 $\mu = 0$， 分散 $\sigma^2 = 1$ （母数は定数）	X：実数， $-\infty < X < \infty$	z 値	大標本 または 母分散が既知 のときの信頼区間の推定，平均の検定 統計量：（標本平均－検定する平均）を，√（分散／標本サイズ）で割ったもの	
t分布 （小標本の分布）	自由度	X：実数， $-\infty < X < \infty$	t 値	小標本 かつ 母分散が未知 のときの信頼区間の推定，平均の検定 統計量：（標本平均－検定する平均）を，√（分散／標本サイズ）で割ったもの	
カイ2乗分布 （誤差の2乗和の分布）	自由度	X：実数， $0 \leq X < \infty$	χ^2 値	Pearsonの χ^2 検定（適合度検定・独立性検定） 統計量：（観測度数－期待度数）の2乗を期待度数で除したものの総和	
F分布 （分散比の分布）	自由度1（分子）， 自由度2（分母）	X：実数， $0 \leqq X < \infty$	F 値	例えば，要因の分散／誤差の分散，信号／ノイズ（S/N比）など，2種類の分散を比率で表したF値（F比）による検定 統計量：分散$_1$ ／ 分散$_2$	

誤差は正規分布になる，これは数が相当量ある事が前提です，境界測量ではその数は少ないことが問題になります。

そこで数が少ない場合の推定，検定に使うのがt分布です，t分布はt検定と平均値信頼区

間の推定に使います，データ数が相当にある場合はχ²分布を使います。t 検定とχ²二乗検定のデータ数の境は自由度（データ数－1）30以下ではt 検定を31以上ではχ²二乗検定を使うとされています。経験上ではt 検定では31個以上でも影響は出ません，χ²二乗検定では20個以上でも影響は出ません，つまりt 検定とχ²二乗検定で差はないようです。

20個～30個の重なる部分は両方の結果を比較して判断すれば良いと思います。質の良いデータでは差が少なくなります。

データの中に異常なデータがないかをt 分布のt 値を使って検定するのがt 検定です，データが正規分布と判断出来るか否かをχ²分布のχ²値を使って検定するのがχ²二乗検定です。

t 分布のt 値は平均値の信頼区間の計算にも使います。χ²値は標準偏差の信頼限界の計算にも使います。

F分布のf 値は二つの測量図を測量してそれぞれの標準偏差から二つの測量図データに有意差があるかないかの判断に使います。差があればどちらかを採用しない，差がなければ両方を同一データとして平均を使うなどと判断していきます，が測量図では有意差がなくても個別に検討しますので実用的な指標かどうかは疑問があります。そのような考え方もあると言おう程度の指標と覚えておけばいいでしょう。

土地家屋調査士がこれらの分布，検定について理解しておく必要があるかと言えば正規分布，t 分布については必要です，他の2つ，χ²二乗分布，F分布についてはどのような時に使われるかを知っておく必要があります。検定では「t 検定」と「χ²二乗検定」については理解しておく必要があります。

土地家屋調査士が測量屋さんと言われる人たちとの違いを出すには最小二乗法と相まってこれらの統計知識が必要です。

データが正規分布に収まっていればそれらのデータを使って境界復元計算することを当たり前に考えなければなりません，限られたデータ数の中でどうやって判断するのか，ここが重要になります，統計でデータの解析をしたうえで最小二乗法を用いて計算をする，その上で法的な判断を加える訳ですが，注意したいのは**統計で許される範囲を超えて法的判断をしてはならないと言うことです。**

具体的には平均値は信頼区間の範囲を超えてはならない，精度は信頼限界を超えてはならない，この範囲を超えた位置に新たに筆界を認定してはならないと言うことです。

えてして，筆界確認が成されたことを理由に筆界位置を変えた地積測量図，地籍図を見ますがこれは行ってはならないことです。ただ誤りということは否定出来ませんがその頻度は極めて少ないと言うことです。

実際には学問的な統計判断を実務にどうやって反映させるかが課題なのです，次に分布に

ついて簡単に触れます，詳細は別途学習してください。

無限のデータを母集団といい，抜き取ったデータを標本と言います，そもそも母集団とはどんなものなのかと言うことです，測量で例えればある条件下，決められた作業基準（作業標準ともいう）の基で得られた成果を母集団といいます。

例えば法17条地図の作成にあたっては5W1H（When（いつ） Where（どこで） Who（誰が） What（何を） Why（なぜ）How（どのように））に添って作業基準を設定した上で得られたデータということです，ある法17条地図の境界点数が2万点あれば母集団は2万点です，この区域内の一区画（一筆）の測量を依頼され測った点が20個であれば標本を20個として考えます。

2万点のデータが正規分布なのだから20個のデータも正規分布になっていると考える，今ある20個のデータは完全な正規分布になっていないけれどもデータ数が増えれば正規分布になると言うのが中心極限定理であり，境界図のデータが正規分布に成っているとすることです。

明治の地租改正図も字単位に作成されましたので一字のデータが母集団という考え方です，地租改正図は市街地，農耕地（郷村地），山林原野の3つに区分けされて異なる基準で作成されましたので同じ字内でも区域が混在していれば分けて母集団が形成されていることに注意しなければなりません。

法17条地図の母集団を同じ作業基準で作成されたとすれば全国のデータを母集団とする考え方もありますが広い意味ではそうなりますが5W1Hで考えれば一区域で考えるべきです。

では，地積測量図はどう考えるかです，地積測量図は一筆毎にしかありません，仮に Who（誰が）を条件にすればA土地家屋調査士が作成した他の地積測量図も含めて母集団とするのか，測量方法，測量機器が変われば当然母集団も変わります，ですから地積測量図には母集団は存在しないけれども架空の母集団が存在するとして，正規分布な母集団からの標本と見る地積測量図のデータも正規分布になっていると考えるのです。

地積測量図を永久保存するのであれば，統計学的見地からいえばその何％かをランダムに抜きとって検査する，その時代のレベルを把握して置くことが法務局，地方法務局に課せられているのではないかと考えられます。

公差と精度（標準偏差）のとらえ方

公差，基準，許容誤差と表現がありますがこれら公差等は最悪の場合，ここまでは許容できる範囲です，「平均的にはこうだ」とか「一般的にはこう考える」という解釈ではありません。

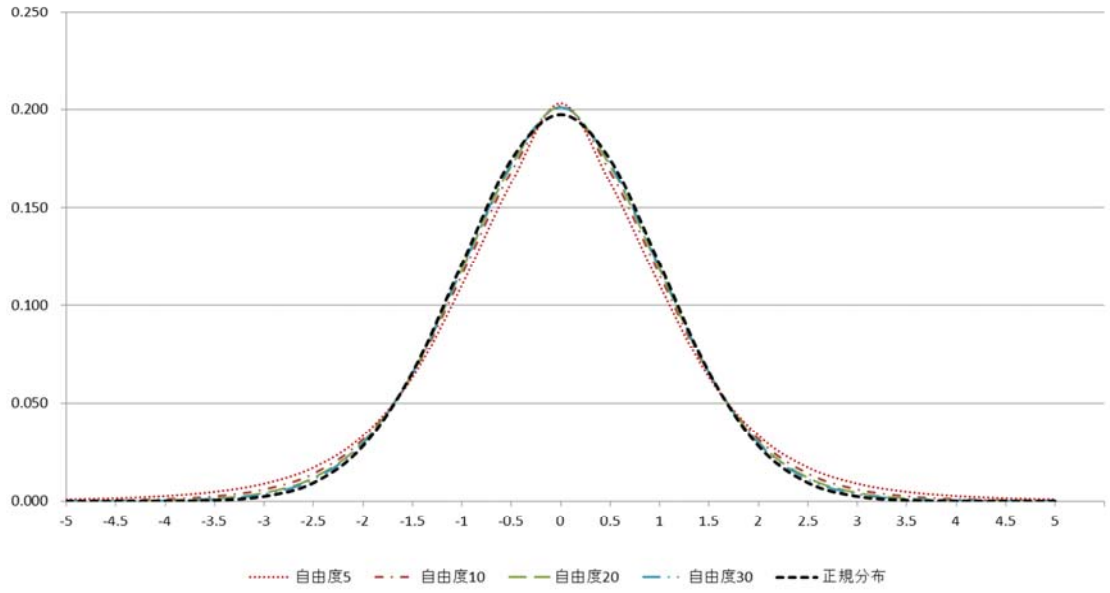
ですから，公差ギリギリの範囲で合格するような測量は通常はありませんししません，公差とは言い方を変えれば誤差の大きさの限界値です，誤差とは既成果実測成果との差です。

分布の形状の概要

t 分布，χ^2二乗分布，f 分布については検定のところで説明しますのでここでは誤差の分布形状について見て下さい。

t 分布

境界(筆界)復元計算で考えるt 分布は t 検定の前提として知っておくべき内容です，統計学ではデータ数によって検定の方法を使い分けます，つまり t 検定では30個以下で使われるとされていますので，境界(筆界)復元では座標変換を使う関係からデータ数を最低でも7個で程度からですから自由度(データ数－1)を 5，10，15，20，30個のグラフを下図にします。



特徴は自由度が増すと正規分布(黒の点線)に近づくこと，自由度が少なくなると釣鐘型の先端が尖って来る(オレンジ色の細点線)，裾があがってくるという特徴があります。

χ^2(カイ)二乗分布

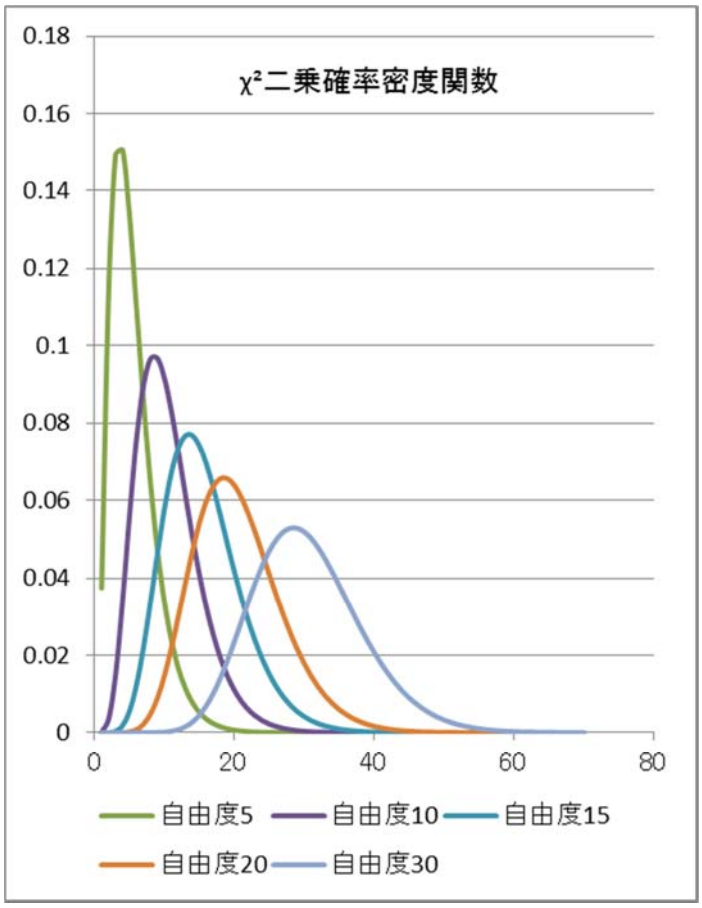
χ^2(カイ)二乗分布の自由度はデータ数ではなく，ヒストグラムで言うところの標準偏差の幅の数－3 の数値です。

ですから，幅を標準偏差の倍数を幾つに取るか，0.5 倍か 0.2 倍の倍数を取るかによって自由度は違って来ます。

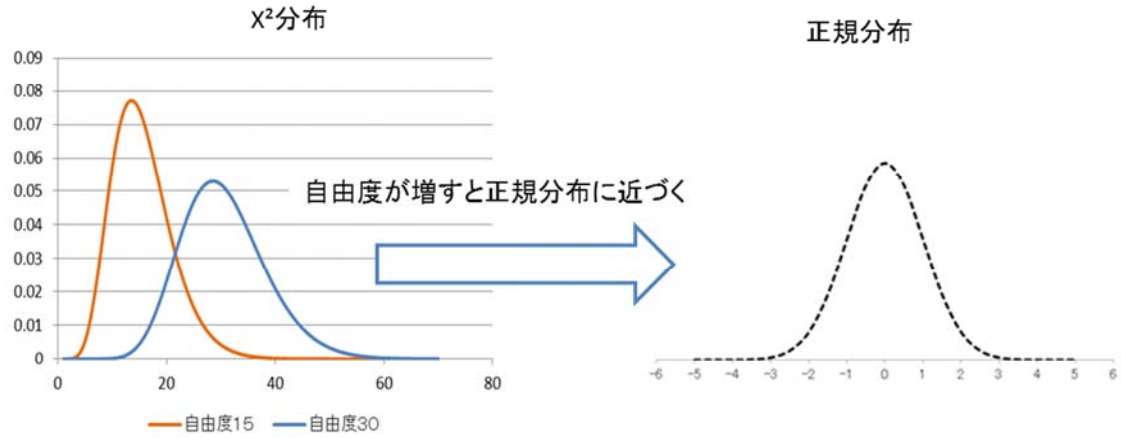
一般的には標準偏差を幅とすればデータは概ね片側5倍の標準偏差に入りますので，0から対象に数えれば10個になり，余裕を見ても12個で自由度は－3ですから9であれば充分と言えます。

χ^2(カイ)二乗検定の精度を高めたい場合は 0.5 標準偏差とすれば20個で－3ですから自由度は17程度でしょうか。

そこで自由度5，10，15，20，30 のグラフを示します。境界(筆界)復元ではデータ数の関係で自由度15付近を使います。

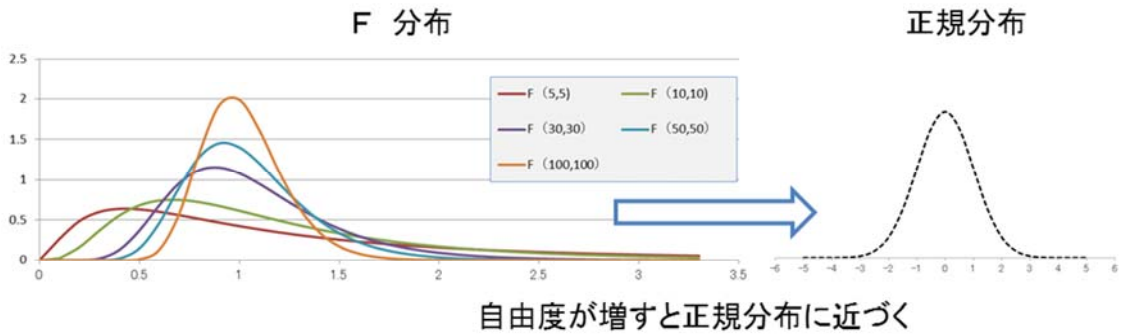


自由度が増すほどに正規分布に近づいていきます。



F分布

F 分布は二つのデータに差があるかどうかの検定に使います F(A 自由度, B自由度)で表します, 自由度はデータ数－1 です。A と B は数が異なっているのが通常ですが数は対, 同じ数にしてあります。詳しくは「F検定」で説明します。



自由度が増すほどに正規分布に近づいていきます。境界（筆界）復元ではデータ数の関係で自由度が5～30の間で判断されることが多いようです。

大まかな説明でしたが，正規分布，t 分布，χ^2二乗分布，F 分布 の四つの分布の形を知っておくことが重要です。

2017／01／20

土地家屋調査士・測量士　小野孝治