確率と最小二乗法による土地境界復元

カイジジョウ　ティー

X^2($\chi 2$）検定、t検定

異常値の有無の確認にt異常値検定
正規分布適合の良否にχ^2（χ２ ） 正規分布適合度検定

2009/6/17作成、2009/6/22、2009/10/01henkanV4.2～に導入、2010/5/21改訂、
2010/07/05改訂

土地の境界復元の基準とする点の選択、準拠点選択にｔ異常値検定（以下ｔ検定という）とχ^2正規分布適合度検定（以下χ^2（χ２ ）検定という）を導入しています。

データのバラツキ、精度、確率または最確値（平均値）を語る訳ですがその前提となるデータは正規分布に従っていることが必要です。

見ただけでは実際にはそのデータが正規分布になっているかどうかは解りませんので統計的検定をおこない確認します、その検定にｔ検定とχ^2（χ２ ）検定を使います。

t検定はデータ中に異常な値がないかの確認に、 χ^2（χ２ ） 検定はデータが正規分布に従っているかの確認に使います、実際にはプログラムに組み込んでありますので検定を意識することはありませんが「こんな方法で計算されている」程度の認識をもっていたほうがよいでしょう。

ｔ検定、 χ^2（χ２ ） 検定といっても様々な種類があります、ここではあくまでもｔ異常値検定とχ^2（χ２ ） 正規分布適合度検定について解説しています。

χ^2（χ２）正規分布適合度検定の式

11点以上で準拠点選択に使う

χ^2（χ２） 検定はχ^2値を期待値と観測値について求めます。

期待値＜観測値　のばあい異常があり

期待値＞観測値　の場合異常が無い　と判断します。

期待値のχ^2値はχ^2分布の限界値表から求めるかエクセルの関数CHIINVから有意水準，自由度を指定して求める。

自由度は級の数－3で、級はヒストグラムの級の数で級の幅の値によって変わります 。

有意水準は1%～5%の間を使いますが統計では一般的に5%が使われます。

観測値のχ^2値は　$=\Sigma((fi-fi*)^2/fi*)$

$(fi-fi*)^2/fi*$　を級毎に求めその合計）

fi は観測度数 （級に当てはまる観測値数）

fi*　は期待度数 （fの合計×級の確率）

検定は統計学上、確立された手法なので特にそのことについての議論はしないで具体的な使い方について説明します。

χ^2（χ２）　検定はデータ数が多い場合にそのデータの組み合わせが正規分布に従うか否かを判断するものです、可能な限り点数が多く正規分布になっている点の組み合わせを探しだします。

それを準拠点に使えば精度の高い復元値が得られるからです。

　χ^2の期待値と観測値の計算式は次の通りです。

観測値のχ^2値は　$=\Sigma((fi-fi*)^2/fi*)$

$((fi-fi*)^2/fi*$　を級毎に求めその合計）

ｆｉ は観測度数　 （級に当てはまる観測値数）

ｆｉ*　は期待度数（ｆｉの合計×確率）　エクセルの関数CHIINVから有意水準，自由度を指定して求めます。

　　χ^2（χ２）　検定のキーポイントはχ^2値（期待値実測値）を求めることです、期待値χ^2値を求めるには自由度を決定しなければなりません、自由度が判れば後の計算は簡単にできます。

標準偏差、χ^2値、ｔ値などの統計指標の計算はエクセルの関数を使ってください、と言いますのは各自で考えた関数を使いますと結果が異なってきますので注意が必要です。

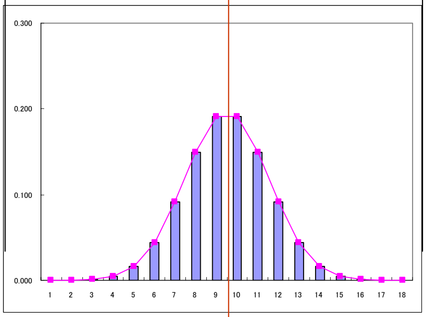
特に測量関係の方は標準偏差を誤って覚えてるケースがありますので必ずエクセル関数でも同じ結果が得られるか確認してください。

次に自由度の計算について説明します。

χ^2（χ２）正規分布適合度検定の自由度

自由度の計算　B

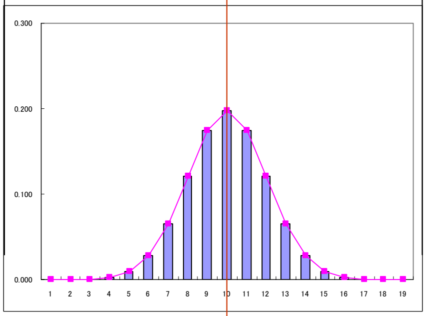
Nが偶数



グラフの境が0

自由度=18−3=15

Nが奇数



グラフの中心が0

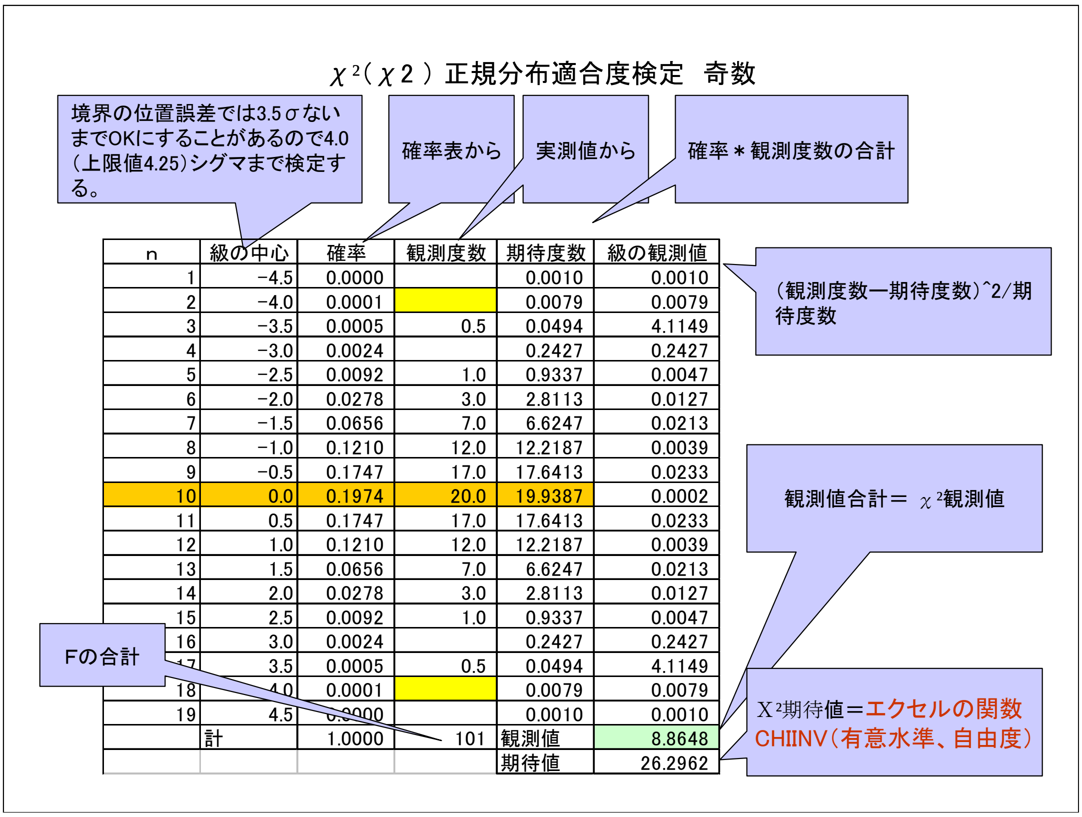
自由度=19−3=16

ｎは級の数です、グラフの中心を0とすればnは奇数になります。

ほぼ同じ考えでグラフの境を0とすればｎは偶数になります。

自由度はｎ－2－1ですから0の位置によって自由度は1つ違ってきます。

0.5σ単位で級を分け、0から4.5σまで級を分けるとグラフの中心を0とすればnは18と19となり自由度は15と16です。



χ²（χ2）正規分布適合度検定　奇数

境界の位置誤差では3.5σないまでOKにすることがあるので4.0（上限値4.25）シグマまで検定する。

確率表から

実測値から

確率＊観測度数の合計

（観測度数ー期待度数）^2/期待度数

n	級の中心	確率	観測度数	期待度数	級の観測値
1	-4.5	0.0000		0.0010	0.0010
2	-4.0	0.0001		0.0079	0.0079
3	-3.5	0.0005	0.5	0.0494	4.1149
4	-3.0	0.0024		0.2427	0.2427
5	-2.5	0.0092	1.0	0.9337	0.0047
6	-2.0	0.0278	3.0	2.8113	0.0127
7	-1.5	0.0656	7.0	6.6247	0.0213
8	-1.0	0.1210	12.0	12.2187	0.0039
9	-0.5	0.1747	17.0	17.6413	0.0233
10	0.0	0.1974	20.0	19.9387	0.0002
11	0.5	0.1747	17.0	17.6413	0.0233
12	1.0	0.1210	12.0	12.2187	0.0039
13	1.5	0.0656	7.0	6.6247	0.0213
14	2.0	0.0278	3.0	2.8113	0.0127
15	2.5	0.0092	1.0	0.9337	0.0047
16	3.0	0.0024		0.2427	0.2427
17	3.5	0.0005	0.5	0.0494	4.1149
18	4.0	0.0001		0.0079	0.0079
19	4.5	0.0000		0.0010	0.0010
	計	1.0000	101	観測値	8.8648
				期待値	26.2962

Fの合計

観測値合計＝ χ²観測値

X²期待値＝エクセルの関数CHIINV（有意水準、自由度）

グラフの中心が0の場合

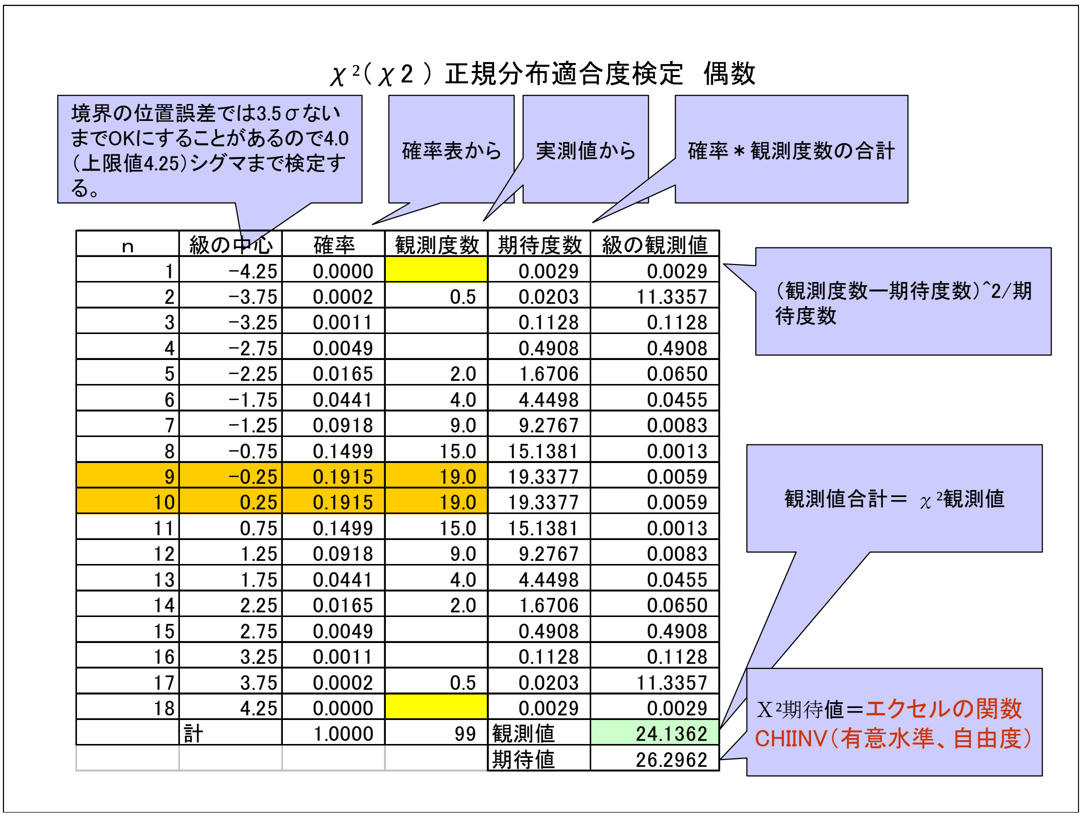
ｎは級をいくつまで計算するかによって決まる、解析精度を上げるため0.5σ単位で計算していく。

ｎは17あれば異常値の解析は出来る、よって自由度は14でよい。有意水準を0.01～0.10の間で影響を見れば自由度は14～17の間で決定されていれば十分である。

境界点のデータはデータ数がもともと少ないので自由度にこだわるメリットはないと考える。

表の場合は由度は17－3＝14　となる。

（ χ²検定体験版を参照）



χ^2（χ2）正規分布適合度検定　偶数

境界の位置誤差では3.5σないまでOKにすることがあるので4.0(上限値4.25)シグマまで検定する。

確率表から

実測値から

確率＊観測度数の合計

n	級の中心	確率	観測度数	期待度数	級の観測値
1	−4.25	0.0000		0.0029	0.0029
2	−3.75	0.0002	0.5	0.0203	11.3357
3	−3.25	0.0011		0.1128	0.1128
4	−2.75	0.0049		0.4908	0.4908
5	−2.25	0.0165	2.0	1.6706	0.0650
6	−1.75	0.0441	4.0	4.4498	0.0455
7	−1.25	0.0918	9.0	9.2767	0.0083
8	−0.75	0.1499	15.0	15.1381	0.0013
9	−0.25	0.1915	19.0	19.3377	0.0059
10	0.25	0.1915	19.0	19.3377	0.0059
11	0.75	0.1499	15.0	15.1381	0.0013
12	1.25	0.0918	9.0	9.2767	0.0083
13	1.75	0.0441	4.0	4.4498	0.0455
14	2.25	0.0165	2.0	1.6706	0.0650
15	2.75	0.0049		0.4908	0.4908
16	3.25	0.0011		0.1128	0.1128
17	3.75	0.0002	0.5	0.0203	11.3357
18	4.25	0.0000		0.0029	0.0029
	計	1.0000	99	観測値	24.1362
				期待値	26.2962

(観測度数ー期待度数)^2/期待度数

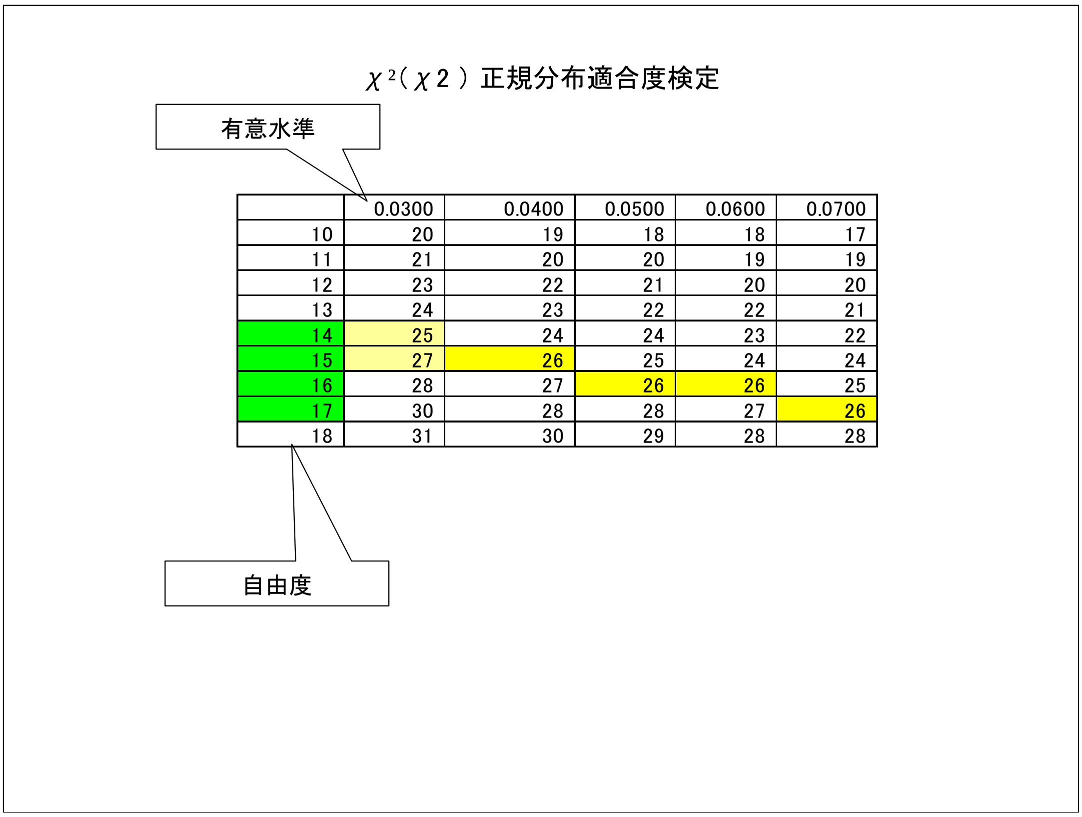
観測値合計＝ χ^2観測値

χ^2期待値＝エクセルの関数CHIINV(有意水準、自由度)

グラフの境が0の場合

表の場合は由度は18−3＝15　となる。

（ χ^2検定体験版を参照）



χ^2(χ2）正規分布適合度検定

有意水準

	0.0300	0.0400	0.0500	0.0600	0.0700
10	20	19	18	18	17
11	21	20	20	19	19
12	23	22	21	20	20
13	24	23	22	22	21
14	25	24	24	23	22
15	27	26	25	24	24
16	28	27	26	26	25
17	30	28	28	27	26
18	31	30	29	28	28

自由度

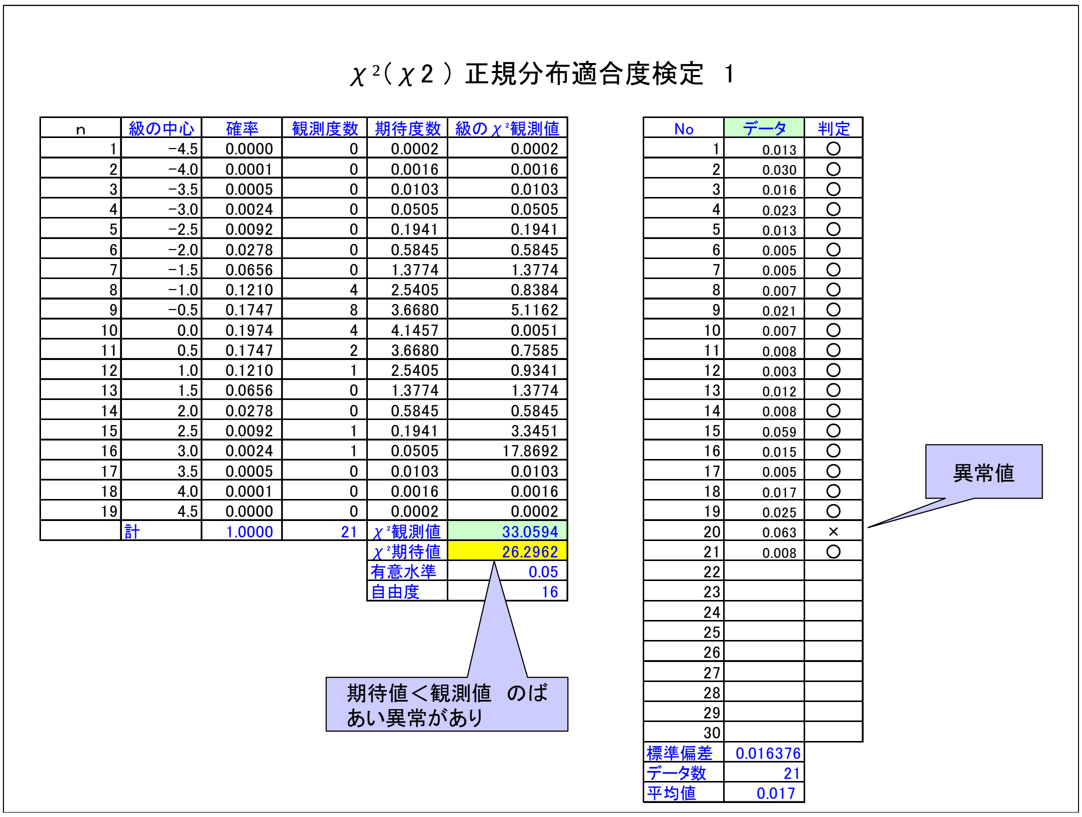
自由度をシビアにしなければならないのかという観点から考えると。

　χ^2値を26と仮定すれば有意水準を0.03～0.07に弾力的に考えれば自由度は14～16の間にあればよいのではないか。

座標変換では本来は円に近い分布の検定であるべきが実際は楕円になっているので自由度が14～16の間の幾つかではなく楕円を解析できるかが問題になる。

これらのことから考えれば自由度は幅を持って考えればよい、実際のデータ検定では14～18の間で準拠の安定した点の組み合わせを選んでいる。

同じ準拠点が選択されれば自由度自体は重要にはならない。



χ^2(χ2) 正規分布適合度検定　1

n	級の中心	確率	観測度数	期待度数	級のχ^2観測値
1	−4.5	0.0000	0	0.0002	0.0002
2	−4.0	0.0001	0	0.0016	0.0016
3	−3.5	0.0005	0	0.0103	0.0103
4	−3.0	0.0024	0	0.0505	0.0505
5	−2.5	0.0092	0	0.1941	0.1941
6	−2.0	0.0278	0	0.5845	0.5845
7	−1.5	0.0656	0	1.3774	1.3774
8	−1.0	0.1210	4	2.5405	0.8384
9	−0.5	0.1747	8	3.6680	5.1162
10	0.0	0.1974	4	4.1457	0.0051
11	0.5	0.1747	2	3.6680	0.7585
12	1.0	0.1210	1	2.5405	0.9341
13	1.5	0.0656	0	1.3774	1.3774
14	2.0	0.0278	0	0.5845	0.5845
15	2.5	0.0092	1	0.1941	3.3451
16	3.0	0.0024	1	0.0505	17.8692
17	3.5	0.0005	0	0.0103	0.0103
18	4.0	0.0001	0	0.0016	0.0016
19	4.5	0.0000	0	0.0002	0.0002
	計	1.0000	21	χ^2観測値	33.0594
				χ^2期待値	26.2962
				有意水準	0.05
				自由度	16

期待値＜観測値　のばあい異常があり

No	データ	判定
1	0.013	○
2	0.030	○
3	0.016	○
4	0.023	○
5	0.013	○
6	0.005	○
7	0.005	○
8	0.007	○
9	0.021	○
10	0.007	○
11	0.008	○
12	0.003	○
13	0.012	○
14	0.008	○
15	0.059	○
16	0.015	○
17	0.005	○
18	0.017	○
19	0.025	○
20	0.063	×
21	0.008	○
22		
23		
24		
25		
26		
27		
28		
29		
30		
標準偏差	0.016376	
データ数	21	
平均値	0.017	

異常値

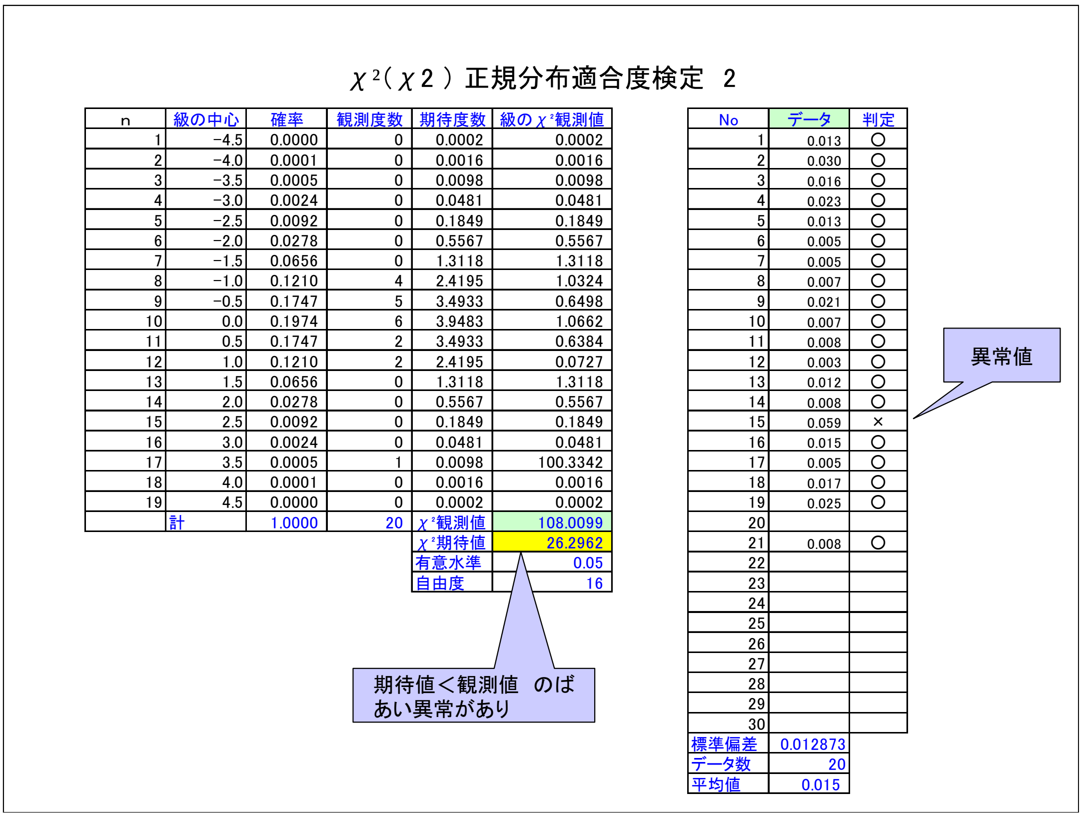
境界を扱うデータはx、yの二変数ですがここでは簡単に一変数のデータについて説明します。

データ（x）を入力。

　χ^2（χ２ ）定体験版を参照してください。

異常値があれば判定に×がつきます、そのデータを消します。

変数減少法と言いましてこの計算を判定に×がなくなるまで繰り返します・・・・1回目の結果　　0.063が異常です。



χ^2（χ2）正規分布適合度検定　2

n	級の中心	確率	観測度数	期待度数	級のχ^2観測値
1	−4.5	0.0000	0	0.0002	0.0002
2	−4.0	0.0001	0	0.0016	0.0016
3	−3.5	0.0005	0	0.0098	0.0098
4	−3.0	0.0024	0	0.0481	0.0481
5	−2.5	0.0092	0	0.1849	0.1849
6	−2.0	0.0278	0	0.5567	0.5567
7	−1.5	0.0656	0	1.3118	1.3118
8	−1.0	0.1210	4	2.4195	1.0324
9	−0.5	0.1747	5	3.4933	0.6498
10	0.0	0.1974	6	3.9483	1.0662
11	0.5	0.1747	2	3.4933	0.6384
12	1.0	0.1210	2	2.4195	0.0727
13	1.5	0.0656	0	1.3118	1.3118
14	2.0	0.0278	0	0.5567	0.5567
15	2.5	0.0092	0	0.1849	0.1849
16	3.0	0.0024	0	0.0481	0.0481
17	3.5	0.0005	1	0.0098	100.3342
18	4.0	0.0001	0	0.0016	0.0016
19	4.5	0.0000	0	0.0002	0.0002
	計	1.0000	20	χ^2観測値	108.0099
				χ^2期待値	26.2962
				有意水準	0.05
				自由度	16

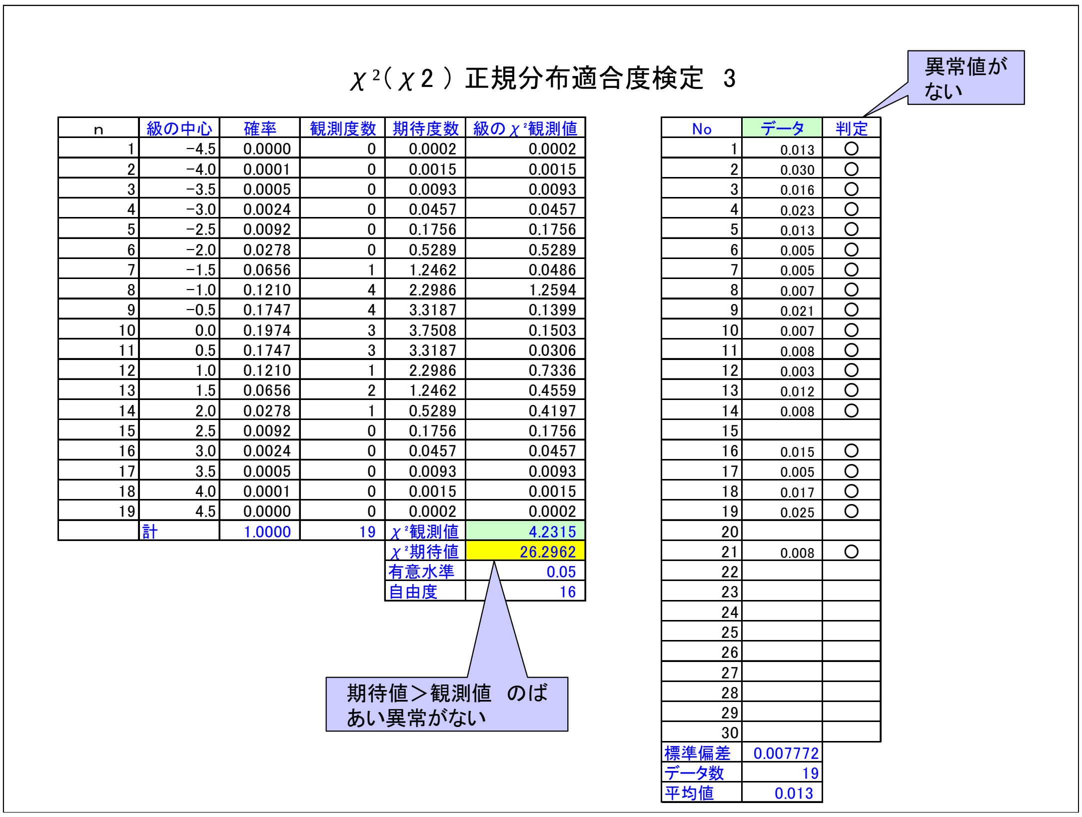
期待値＜観測値　のばあい異常があり

No	データ	判定
1	0.013	○
2	0.030	○
3	0.016	○
4	0.023	○
5	0.013	○
6	0.005	○
7	0.005	○
8	0.007	○
9	0.021	○
10	0.007	○
11	0.008	○
12	0.003	○
13	0.012	○
14	0.008	○
15	0.059	×
16	0.015	○
17	0.005	○
18	0.017	○
19	0.025	○
20		
21	0.008	○
22		
23		
24		
25		
26		
27		
28		
29		
30		
標準偏差	0.012873	
データ数	20	
平均値	0.015	

異常値

異常値があれば判定に×がつきます、そのデータをけします。

この計算を判定に×がなくなるまで繰り返します・・・・2回目の結果　　0.059が異常です。



χ^2（χ2）正規分布適合度検定　3

n	級の中心	確率	観測度数	期待度数	級のχ^2観測値
1	−4.5	0.0000	0	0.0002	0.0002
2	−4.0	0.0001	0	0.0015	0.0015
3	−3.5	0.0005	0	0.0093	0.0093
4	−3.0	0.0024	0	0.0457	0.0457
5	−2.5	0.0092	0	0.1756	0.1756
6	−2.0	0.0278	0	0.5289	0.5289
7	−1.5	0.0656	1	1.2462	0.0486
8	−1.0	0.1210	4	2.2986	1.2594
9	−0.5	0.1747	4	3.3187	0.1399
10	0.0	0.1974	3	3.7508	0.1503
11	0.5	0.1747	3	3.3187	0.0306
12	1.0	0.1210	1	2.2986	0.7336
13	1.5	0.0656	2	1.2462	0.4559
14	2.0	0.0278	1	0.5289	0.4197
15	2.5	0.0092	0	0.1756	0.1756
16	3.0	0.0024	0	0.0457	0.0457
17	3.5	0.0005	0	0.0093	0.0093
18	4.0	0.0001	0	0.0015	0.0015
19	4.5	0.0000	0	0.0002	0.0002
	計	1.0000	19	χ^2観測値	4.2315
				χ^2期待値	26.2962
				有意水準	0.05
				自由度	16

期待値＞観測値　のばあい異常がない

No	データ	判定
1	0.013	○
2	0.030	○
3	0.016	○
4	0.023	○
5	0.013	○
6	0.005	○
7	0.005	○
8	0.007	○
9	0.021	○
10	0.007	○
11	0.008	○
12	0.003	○
13	0.012	○
14	0.008	○
15		
16	0.015	○
17	0.005	○
18	0.017	○
19	0.025	○
20		
21	0.008	○
22		
23		
24		
25		
26		
27		
28		
29		
30		
標準偏差	0.007772	
データ数	19	
平均値	0.013	

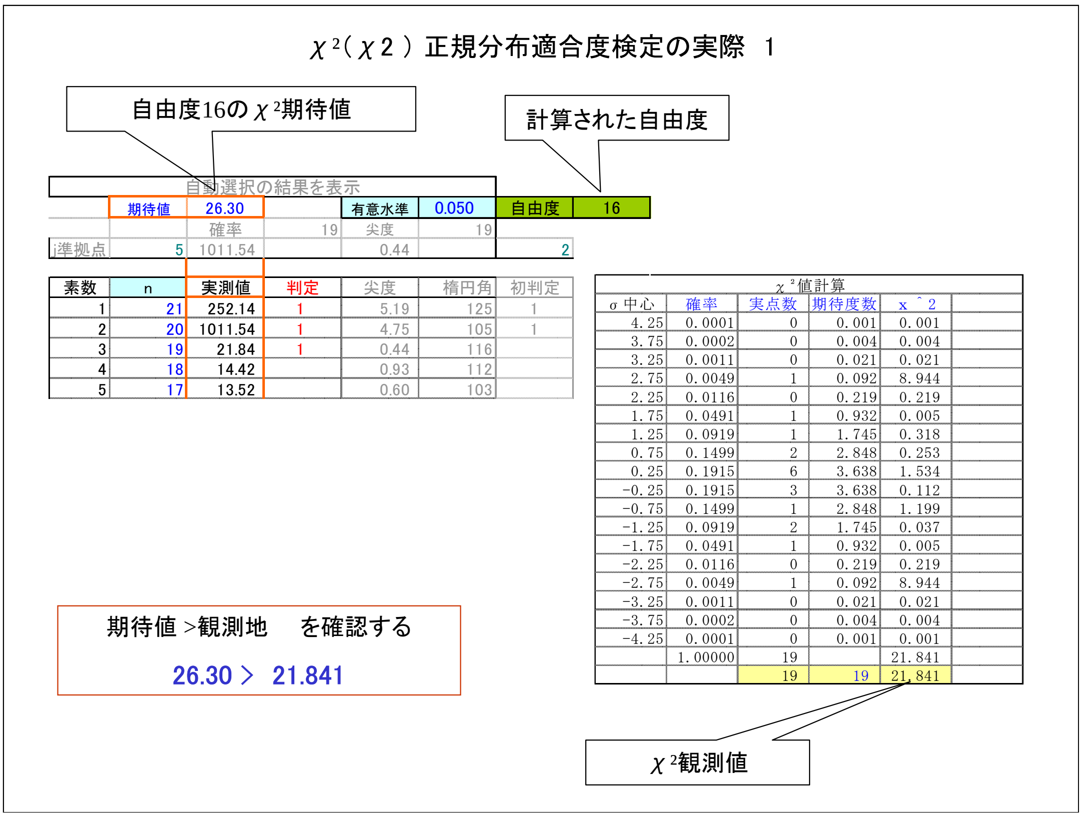
異常値がない

異常値があれば判定に×がつきます、そのデータをけします。

二回目のデータを除いた段階で異常値はありませんのでここで終了します。

このように変数減少法によって異常値を除いていきます。

その結果0.063と0.059が異常な値と判断します、実際には有意水準、自由度によって変化しますので最終的に結果を踏まえて判断することになります。



χ^2（χ2）正規分布適合度検定の実際　1

自由度16のχ^2期待値

計算された自由度

自動選択の結果を表示							
	期待値	26.30		有意水準	0.050	自由度	16
		確率	19	尖度	19		
j準拠点	5	1011.54		0.44		2	

素数	n	実測値	判定	尖度	楕円角	初判定
1	21	252.14	1	5.19	125	1
2	20	1011.54	1	4.75	105	1
3	19	21.84	1	0.44	116	
4	18	14.42		0.93	112	
5	17	13.52		0.60	103	

期待値 >観測地　を確認する

26.30 >　21.841

χ^2値計算

σ中心	確率	実点数	期待度数	x＾2
4.25	0.0001	0	0.001	0.001
3.75	0.0002	0	0.004	0.004
3.25	0.0011	0	0.021	0.021
2.75	0.0049	1	0.092	8.944
2.25	0.0116	0	0.219	0.219
1.75	0.0491	1	0.932	0.005
1.25	0.0919	1	1.745	0.318
0.75	0.1499	2	2.848	0.253
0.25	0.1915	6	3.638	1.534
-0.25	0.1915	3	3.638	0.112
-0.75	0.1499	1	2.848	1.199
-1.25	0.0919	2	1.745	0.037
-1.75	0.0491	1	0.932	0.005
-2.25	0.0116	0	0.219	0.219
-2.75	0.0049	1	0.092	8.944
-3.25	0.0011	0	0.021	0.021
-3.75	0.0002	0	0.004	0.004
-4.25	0.0001	0	0.001	0.001
	1.00000	19		21.841
		19	19	21.841

χ^2観測値

実際の計算では自由度から求めたχ^2期待値＞χ^2観測値　の関係を確認すれば検定の終了です。

有意水準は5％で行っています。

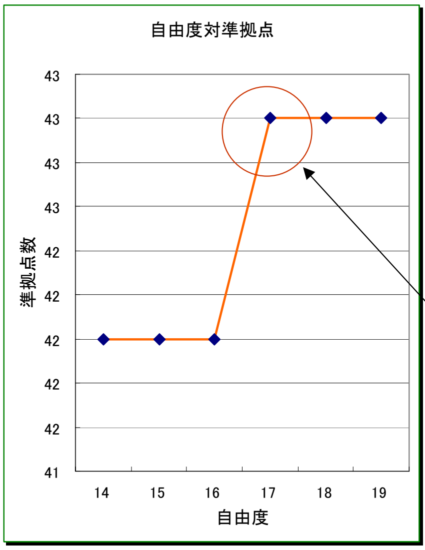
次に説明するｔ検定もそうですが検定を測量データに使うことは出来ないのではと考えられておりました、理論としてはあったのですが実際に使われた実例は見たことがありません。

当webでは過去4年半の（2005年～2009年）データ解析の結果を検証してからマトメたもので教科書の解説に沿って作成したものではありません。

χ^2(χ２）正規分布適合度検定の実際　2

成立の例　自由度対準拠点の関係を調べてみる。

自由度は14、15，16で
どれかで決定されるようで
す、過去一年のデータで
も自由度が17のデータが
ありましたが17の場合は
安定性に欠けるばあいで
この時は16で決定すれ
ばいいようです。



自由度17(nが奇数と
判断されている)

自由度14〜19の間で準拠点選択が安定する点を探し、自由度を逆算する方法をとっています。

　χ^2検定で自由度は幾つになるかが重要です、理論的には14か15（級が１７か１８で）ですが14〜16の間にあるものと推定して、グラフのように安定する組み合わせを逆に探し出し自由度を求めてから準拠点選択を決定する手法をとっています。

決定された自由度が推定される自由度14〜16の間にあれば復元精度には問題ありません。

自由度の決定はデータ数が多ければ問題はなく十数点の組み合わせで精度の悪いデータ（準拠点の組み合わせで標準偏差が大きく変化するデータ）に注意が必要です。

t異常値検定の式

10点以下で異常点の除去と準拠点選択に使う

t検定はt値を期待値と観測値について求め

期待値＜観測値　のばあい異常がある

期待値＞観測値　の場合異常が無い　と判断します。

期待値のｔ値はｔ分布の限界値表から求めるかエクセルの関数TINVから有意水準，自由度を指定して求めます。

有意水準は1%～5%の間を使いますが統計では一般的に5%が使われます。

自由度は観測数（ｎ）－2です。

観測値のt値は $=((\sqrt{n-2})\times\tau)/\sqrt{(n-1-\tau^2)}$

$\tau=(xk-xm)/\sigma$

xkは判定する観測値、xmは観測値の平均値、σはｎ（観測数）を使った標準偏差で求めます。

ｔ検定はｔ値を期待値と観測値について求め　期待値＜観測値　のばあい異常がある、期待値＞観測値　の場合異常が無いと判断するものです。

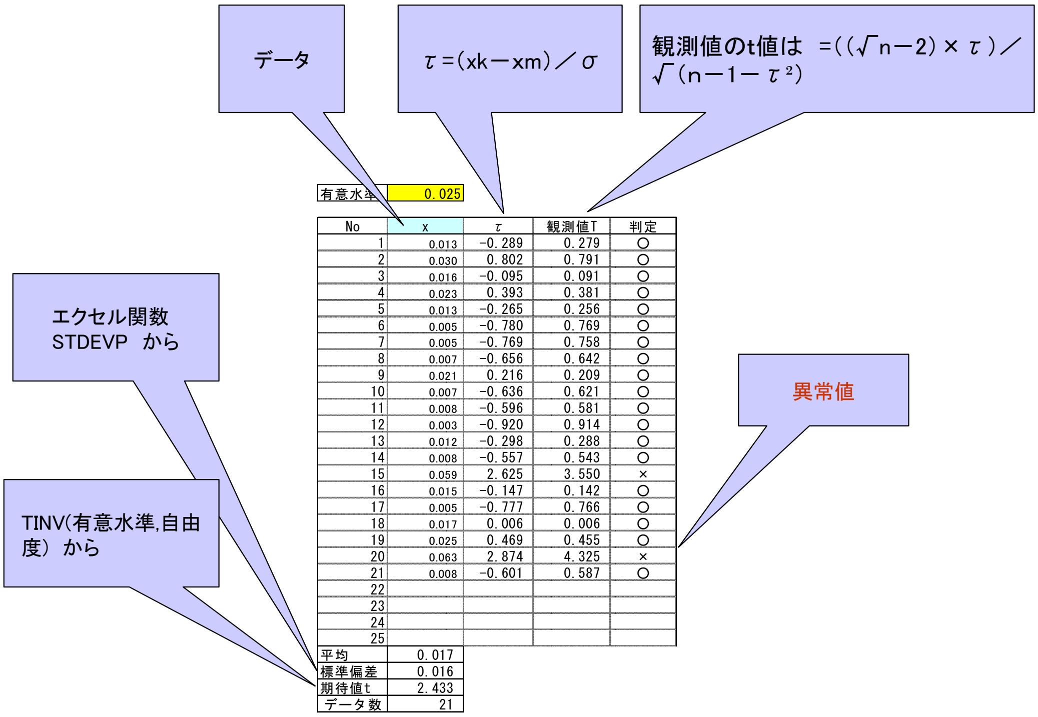
期待値のｔ値は観測数からｔ分布の限界値表から求めるかエクセルのTINV関数からTINV（有意水準，自由度）で求めます、有意水準は1%～5%の間を使いますが統計では一般的に5%が使われます、自由度は観測数－2です。

観測値のｔ値は　$=((\sqrt{n-2})\times\tau)/\sqrt{(n-1-\tau^2)}$

$\tau=(xk-xm)/\sigma$　　xkは判定する観測値、xmは観測値の平均値、σはｎを使った標準偏差　で求めます。

この式を埋め込んでエクセルのプログラムを作って計算してみます。

t異常値検定　1



データ

τ＝(xk－xm)／σ

観測値のt値は　=((√n－2)×τ)／√(n－1－τ^2)

エクセル関数
STDEVP　から

TINV(有意水準,自由度）から

異常値

有意水準	0.025

No	x	τ	観測値T	判定
1	0.013	−0.289	0.279	○
2	0.030	0.802	0.791	○
3	0.016	−0.095	0.091	○
4	0.023	0.393	0.381	○
5	0.013	−0.265	0.256	○
6	0.005	−0.780	0.769	○
7	0.005	−0.769	0.758	○
8	0.007	−0.656	0.642	○
9	0.021	0.216	0.209	○
10	0.007	−0.636	0.621	○
11	0.008	−0.596	0.581	○
12	0.003	−0.920	0.914	○
13	0.012	−0.298	0.288	○
14	0.008	−0.557	0.543	○
15	0.059	2.625	3.550	×
16	0.015	−0.147	0.142	○
17	0.005	−0.777	0.766	○
18	0.017	0.006	0.006	○
19	0.025	0.469	0.455	○
20	0.063	2.874	4.325	×
21	0.008	−0.601	0.587	○
22				
23				
24				
25				

平均	0.017
標準偏差	0.016
期待値t	2.433
データ数	21

境界を扱うデータはx、yの二変数ですがここでは簡単に一変数のデータについて説明します。

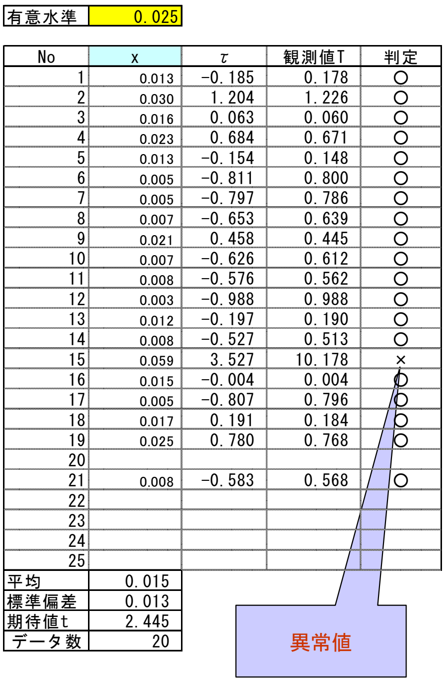
①始めに得られたデータを大きい順に並べます、そのままでもかまいませんが並べたほうが後で結果が分かりやすい。

②データ（x）を入力。

τ、観測値τ、期待値ｔの式をセルにはめ込んでおけば計算されます（ｔ検定体験版を参照）。

この計算を判定に×がなくなるまで繰り返します・・・・1回目の結果　　0.063が異常

t異常値検定 2

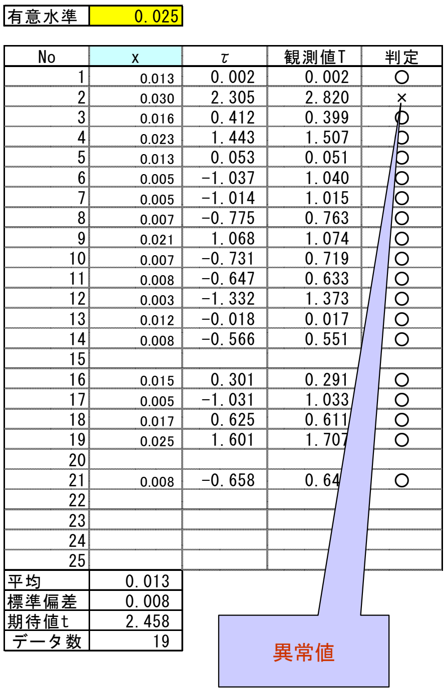


有意水準	0.025

No	x	τ	観測値T	判定
1	0.013	-0.185	0.178	○
2	0.030	1.204	1.226	○
3	0.016	0.063	0.060	○
4	0.023	0.684	0.671	○
5	0.013	-0.154	0.148	○
6	0.005	-0.811	0.800	○
7	0.005	-0.797	0.786	○
8	0.007	-0.653	0.639	○
9	0.021	0.458	0.445	○
10	0.007	-0.626	0.612	○
11	0.008	-0.576	0.562	○
12	0.003	-0.988	0.988	○
13	0.012	-0.197	0.190	○
14	0.008	-0.527	0.513	○
15	0.059	3.527	10.178	×
16	0.015	-0.004	0.004	○
17	0.005	-0.807	0.796	○
18	0.017	0.191	0.184	○
19	0.025	0.780	0.768	○
20				
21	0.008	-0.583	0.568	○
22				
23				
24				
25				

平均	0.015
標準偏差	0.013
期待値t	2.445
データ数	20

異常値



有意水準	0.025

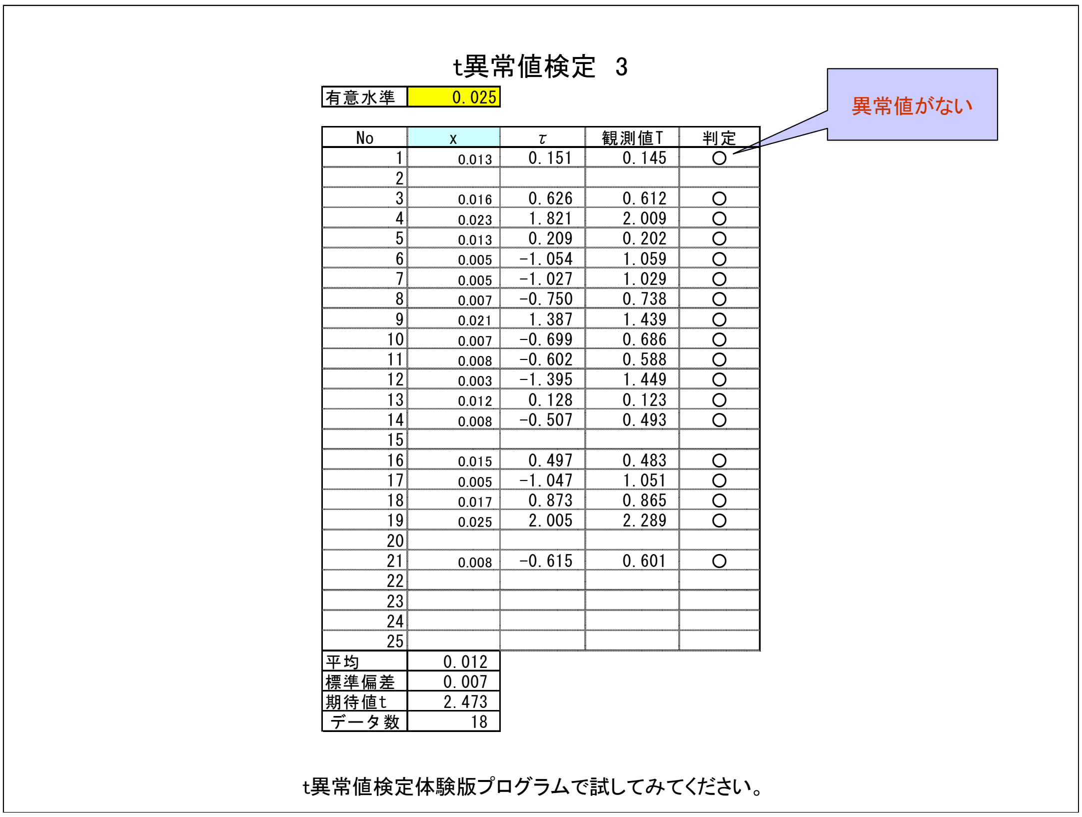
No	x	τ	観測値T	判定
1	0.013	0.002	0.002	○
2	0.030	2.305	2.820	×
3	0.016	0.412	0.399	○
4	0.023	1.443	1.507	○
5	0.013	0.053	0.051	○
6	0.005	-1.037	1.040	○
7	0.005	-1.014	1.015	○
8	0.007	-0.775	0.763	○
9	0.021	1.068	1.074	○
10	0.007	-0.731	0.719	○
11	0.008	-0.647	0.633	○
12	0.003	-1.332	1.373	○
13	0.012	-0.018	0.017	○
14	0.008	-0.566	0.551	○
15				
16	0.015	0.301	0.291	○
17	0.005	-1.031	1.033	○
18	0.017	0.625	0.611	○
19	0.025	1.601	1.707	○
20				
21	0.008	-0.658	0.64	○
22				
23				
24				
25				

平均	0.013
標準偏差	0.008
期待値t	2.458
データ数	19

異常値

この計算を判定に×がなくなるまで繰り返します・・・・2回目の結果　　0.059が異常

この計算を判定に×がなくなるまで繰り返します・・・・3回目の結果　　0.030が異常



t異常値検定　3

有意水準	0. 025

No	x	τ	観測値T	判定
1	0.013	0. 151	0. 145	○
2				
3	0.016	0. 626	0. 612	○
4	0.023	1. 821	2. 009	○
5	0.013	0. 209	0. 202	○
6	0.005	-1. 054	1. 059	○
7	0.005	-1. 027	1. 029	○
8	0.007	-0. 750	0. 738	○
9	0.021	1. 387	1. 439	○
10	0.007	-0. 699	0. 686	○
11	0.008	-0. 602	0. 588	○
12	0.003	-1. 395	1. 449	○
13	0.012	0. 128	0. 123	○
14	0.008	-0. 507	0. 493	○
15				
16	0.015	0. 497	0. 483	○
17	0.005	-1. 047	1. 051	○
18	0.017	0. 873	0. 865	○
19	0.025	2. 005	2. 289	○
20				
21	0.008	-0. 615	0. 601	○
22				
23				
24				
25				

平均	0. 012
標準偏差	0. 007
期待値t	2. 473
データ数	18

異常値がない

t異常値検定体験版プログラムで試してみてください。

この計算を判定に×がなくなるまで繰り返します・・・・4回目の結果　　異常なし

結果は0. 063、0. 059，0. 030が異常値の判断です。

この方法を変数減少法といいます。

変数減少法までは統計の解説本には書いてないです、実務ではここまで詰めないと使えません。

t異常値検定の実際　1

このような図面の成果対観測値を検定する場合「データ」に何の数値をもってくるか

図面値			実測値		
点名	X	Y	点名	X	Y
Z9	100	63.64	G9	-491.882	137.259
Z10	113.726	63.64	G10	-485.386	125.22
Z12	115.262	55.31	G12	-491.804	119.784
Z15	113.697	46.277	G15	-500.303	116.625
Z16	114.758	37.505	G16	-507.474	111.347
Z17	116.273	35.622	G17	-508.807	110.292
Z18	98.855	34.54	G18	-517.34	123.707
Z19	99.215	45.441	G19	-507.945	129
Z20	99.622	54.649	G20	-499.901	133.116

変換された座標値		
点名	X	Y
aZ9	-491.870	137.255
aZ10	-485.133	125.230
aZ12	-491.657	119.746
aZ15	-500.317	116.630
aZ16	-507.461	111.343
aZ17	-508.362	109.080
aZ18	-517.857	123.801
aZ19	-508.156	128.902
aZ20	-499.911	133.120

図面値対実測値のデータは基本的に測った次元が違いますのでこのデータを図に歪みが無い場合はヘルマート変換で図に歪みがある場合はアフィン変換で変換値をもとめます。

多角点が異なる場合、座標軸が異なる場合、同じ多角点を使用している場合でも変換は行います。

この実測値と変換値の差から標準偏差を変数減少法によるフルカウント法（総当り法）によっ計算します。

標準偏差の大きい順に点名と標準偏差の差を求め、この値をデータ差としてｔ検定を行えば異常値が判断できます。

実際には座標値で図面値対観測値の相対的な関係を検証して異常な点を除くことが目的です。

測量データの場合x、yに相関関係がありません、この場合の座標データを解析をする、簡単にいえば有るデータ群から異常なデータを除く方法について説明します。

図面値と現時点の実測値の検定になりますので先のスライドの「データ差」欄に入れる数値を求めなければなりません。

図面値対実測値のデータは基本的に測った次元が違います、このデータに図に歪みが無い場合はヘルマート変換で図に歪みがある場合はアフィン変換で変換値をもとめます。

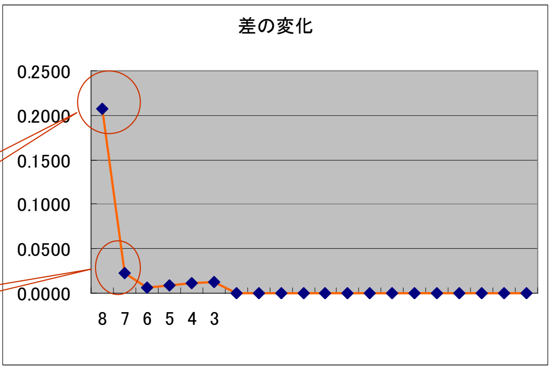
多角点が異なる場合、座標軸が異なる場合、同じ多角点を使用している場合でも変換は行います。

この実測値と変換値の差から標準偏差を変数減少法によるフルカウント法（総当り法）によっ計算します。

標準偏差の大きい順に点名と標準偏差、標準偏差の差を求め、この値をデータ差としてｔ検定を行えば異常値が判断できます。

t異常値検定の実際　2

異常値	残点数	除番号	標準偏差	差
	9		0.269	
×	8	G17	0.062	0.207
×	7	G18	0.039	0.023
	6	G19	0.032	0.007
	5	G15	0.024	0.009
	4	G10	0.012	0.011
	3	G9	0.000	0.012



異常な点

異常な点
グラフでは判りにくい

標準偏差を計算してその値を使えば異常な点の検定ができます。

標準偏差の計算はフルカウント法（総当り法）によって標準偏差の値を大きくしている順番にその点を除いていく方法をとります。

実際のプログラムではさらに工夫を加えて標準偏差の差を検定している。

様々なデータを検証しましたが標準偏差の差を捉えるのが最も適切な復元値が得られることから標準偏差の差を使っています。

ｔ検定の方法は基本的にデータ数が10点以下と少ない場合に使いますが求める変換値がヘルマートの場合7点、アフィンで8点以下で適切は判断ができない場合があります。

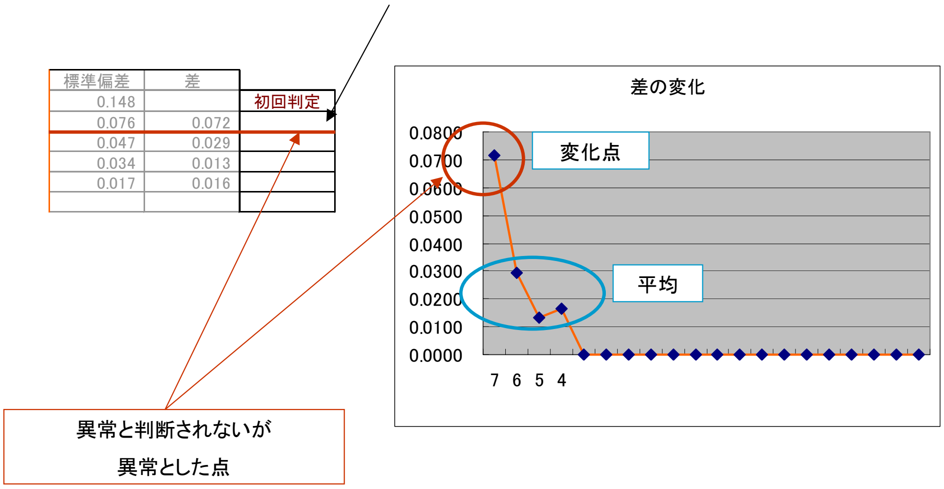
その注意点に付いては次に説明します。

先のχ²検定とｔ検定でデータ数（11点以上で）に関係なく同じ結果が得られるのだろうかと疑問を持つことがあると思います、これについては元のデータが正規分布に近いほど同じ結果が得られます。

t異常値検定の実際　3

データが少ないとt検定では正確に判断できないことがある（n=8、アフィンフルカウントのデータ例）

T検定で異常であればこの欄に“×”が入る。



標準偏差	差	初回判定
0.148		
0.076	0.072	
0.047	0.029	
0.034	0.013	
0.017	0.016	

異常と判断されないが

異常とした点

変化の大きい点を境にその前の点の値と変化後の数点の平均が約2.5倍～3.0倍を超えれば変化点以前の点は異常点と判断できる。

データが正規分布になっているかどうかの判断が難しいので異常値を検定して除けば充分であるがデータがヘルマートで７点以下、アフィンで８点以下とさらに少ないとｔ検定で判断できないことがあります。

この場合は差のグラフを見て判断するすることになります。

この時の判断の基準は概ね変化の大きい点を境にその前の点の値と変化後の数点の平均が約2.5倍～3.0倍を超えれば変化点以前の点は異常点と判断できます。

ただし、準拠点の配点バランス（準拠点が特定の箇所にかたまらないなど）も考慮して行う必要があります、復元値の変化等も考慮して配点バランスを崩さないことが必要です。

土地家屋調査士が扱う筆界のデータは点数が少ないことが多いのでこういった判断は頻繁に出てきます。