χ^2 分布理解のための Excel によるシミュレーション[1)]
―χ^2 値の定義と分布図との関連性について―

門田　幸太郎[i]

χ^2分布の定義から確率分布図を導き出す過程を，EXCELによるシミュレーションを使って，具体的な事象として擬似的に体験することによって，χ^2 分布を直観的に理解することができるようにすることが本稿の目的である。まず，χ^2 検定を利用法とその論理を概説し，χ^2 値の定義について述べる。次に正規分布の確率密度関数を取り上げ，EXCEL の RAND 関数を利用して正規分布での標本抽出法を述べる。その結果，得られた度数分布に対して，FREQUENCY 関数を使って度数分布表の作成し，それに基づいて χ^2 分布図の作成をする。

キーワード： 標準正規分布，χ^2（カイ２乗）分布，χ^2 関数，χ^2 検定，エクセル関数，FREQUENCY 関数，確率密度関数，エクセルによるシミュレーション，VLOOKUP 関数

1．χ^2 検定の利用

χ^2 検定は，名義尺度でも利用可能な検定方法として，きわめて利用頻度の高いものである。この検定方法の根拠となるものが，χ^2 分布である。χ^2 分布は変数の自由度により分布が異なる。自由度が１から６までの χ^2 値の分布曲線をえがくと Figure 1. のようになる。この図からも推測されるように，χ^2 分布は自由度が大きくなればなるほど，その形は正規分布曲線に近づくことになる。一般的には，自由度が30より大きくなれば近似的に正規分布するとされている。

まず，χ^2 分布を利用した χ^2 検定が，得られた度数データをどのような論理にしたがって，有意差検定に利用するのかを概説してみる。χ^2 検定では，実験条件間や被験者群間に差がないという帰無仮説の下で，与えられたデータがどのような生起確率 χ^2 値をもっているかが求められる。その得られた χ^2 値が日常的に起こりうる充分な大きさなのか，まれにしか起こらない稀少な出来事なのかが判断される。一般には，起こりうる大きさとは95％を目処にしている，したがって，稀少な出来事は５％水準とされる。５％水準の確率領域を棄却域と呼び，所与の χ^2 値が棄却域に入るか入らないかが問われる。もし，棄却域に入るならば，帰無仮説は棄却される。ただし，ここで，本当は帰無仮説が正しいにもかかわらずこれを採択しないという間違い（第１種の誤り）を犯す可能性（有意水準または危険率）が設定される。このように，検定は一定の保留条件の下で帰無仮説が否定されて，実験条件間や被験者群間における変数の発生頻度である度数に，統計的に意味の

ｉ　立命館大学産業社会学部教授

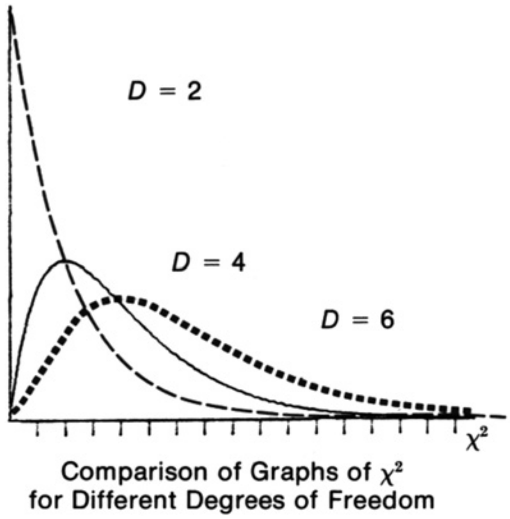

Figure 1. Chi-square distribution of various degrees of freedom.
（Gilbert, 1981）

ある差，すなわち有意差があるか否かが問われる。

Figure 2. は有意水準が α で，自由度が n の場合の χ^2 分布が示されている。この図で黒塗りにされている領域の左端の横軸での値がデータから求められた χ^2 値，$\chi^2_n(\alpha)$ である。この図をもとに，検定の論理をたどってみる。もし得られた $\chi^2_n(\alpha)$ が5％水準，たとえば自由度が5の場合なら11.070となるが，それより大きい場合に，きわめて起こりにくい現象が生じたと考える。そして，このような結果が得られたのは，実験条件間や被験者群間に差がない，つまり同一母集団から抽出されたものという帰無仮説の前提が誤っていたと判断され，その結果，得られたデータは同一母集団からではなく，それぞれ異なる母集団から抽出されたもの，すなわち，実験条件間や被験者群間に有意差があるとされる。もし得られた $\chi^2_n(\alpha)$ が5％水準より小さい場合は，帰無仮説のもと日常的によくみられる現象が生じたに過ぎないと考えて，得られた $\chi^2_n(\alpha)$ に統計的には特段の意味はなく誤差の範囲だとされて帰無仮説は否定されないことになる。もとより，研究者としては帰無仮説が否定

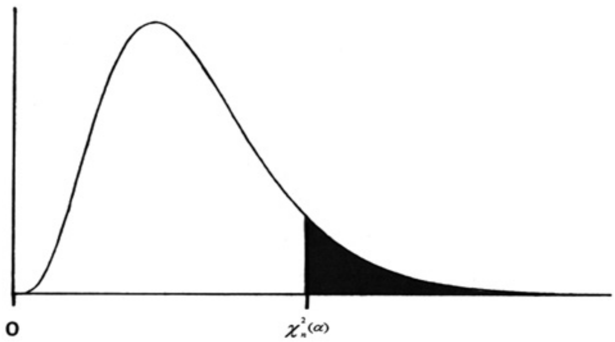

Figure 2. Chi-square distribution of *n* degree of freedom and α significant level.
（Gilbert, 1981 より改変）

されることを期待して研究計画を立てるわけである。以上がχ^2検定の論理であるが，この例でも示されているように，検定のもとになるχ^2分布がなぜこのような分布を取るのかということを理解することが，χ^2検定の理解に欠かせないものとなる。

2．χ^2値の定義

χ^2値の求め方としては，以下のような説明が一般的である。

「確率変数Xが正規分布$N(\mu, \sigma^2)$にしたがっているとき独立に抽出されたn個の標本x_1，x_2，…，x_nから求められた統計量

$$\chi^2 = \sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2$$

は自由度$\phi=n$のχ^2分布にしたがう。」[1]

このχ^2の定義式から確率分布図に至るまでの過程は，少なくとも初学者にとっては，かなり理解しがたいものといえる。この過程を説明するのが本稿の目的である。その手段としてEXCELによるシミュレーションを用いることにする。定義式でχ^2値を求めるときの前提条件として「確率変数Xが正規分布$N(\mu, \sigma^2)$にしたがっているとき」というものがある。そこで，まず，正規分布する確率変数Xを作成することが求められる。

3．正規分布

平均が０，分散が１と標準化された変数が正規分布する場合，標準正規確率分布関数は$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$で表される。Table 1.にその一部を示すように，A2セルに－5.00を入力し，A3セルに－4.90を入力する。その後，A2セルで左クリックしA3セルまでドラッグしてリリースする。これにより，A2セルとA3セルとがフォーカスされた状態になる。このフォーカスされた範囲をA102セルまで広げると，A4セルに－4.80が入力され，以下順次0.1を加えた値がA102セルの5.00まで自動的に入力することができる。

次に，各0.1刻みの中間点を求める。例えば，A2セルの－5.00とA3セルの－4.90の中間点をB3セルに求めるようにする。そのため，B3セルには数式「＝AVERAGE(A2,A3)」を入力する。一般に，AVERAGE（値１，値２，…）は値１，値２，…の平均を求めるためのEXCEL関数である。こうすることによって，0.1単位の幅で相対確率を求めるための中心点が得られる。これが確率変数Xの値となる。

各階級の確率，すなわち相対確率は，Figure 3.の横線の網掛け部分と左上の黒塗りの部分を合わせた面積で求められるが，ここでは，近似値として，横線の網掛け部分と右上の斜線の部分を合わせた長方形面積で求

1　母平均ではなく，標本平均を採用する場合は，以下のように定義される。
μを標本平均$\bar{x}$としたとき，

$$\chi^2 = \sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma}\right)^2$$

は自由度$\phi=n-1$のχ^2分布にしたがう。

本稿においては，母平均を採用する場合について取り上げる。

められる。つまり，黒塗り部分と斜線部分は近似的に相殺されると考える。中間点における確率密度関数の値は，具体的にはこの長方形の高さを意味していることになる。そのため，階級ごとの相対確率を求めるには，その幅である0.1を掛けてやる必要がある。

中間点を確率変数の値として，それに対する確率密度，つまり長方形の高さを求める関数 EXP(－B3*B3/2) SQRT(2*PI()) をC3セルに入力する。これは，$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ を求めるものである。EXP関数はEXP(a) という形式で用いられ，自然対数 e のベキ乗を求める EXCEL 関数である。EXP 定義という関数は自然対数（2.7182818…）を意味し，EXP(a) は e^{a} を意味する。EXP(－B3*B3/2) は $e^{-\frac{x^2}{2}}$ の x に B3 セルの値を入れることを意味する。分母のSQRT(2*PI()) のSQRT( ) は（ ）内の値の平方根を求める関数である。（ ）内のPI( ) は円周率の π を意味する。一般的に，EXCEL 関数の（ ）内には，上の EXP(a) の a にあたる引数を必要とするが，PI( ) の場合は，引引数を必要としない。

Table 1. Usage of AVERAGE function in B3 cell.

B3 =AVERAGE(A2,A3)

	A	B	C	D	E	F
1	階級	×	確率密度	確率	累積確率	個数
2	−5.00			0.0000	0.0000	0
3	−4.90	−4.95	0.0000	0.0000	0.0000	0
4	−4.80	−4.85	0.0000	0.0000	0.0000	0
5	−4.70	−4.75	0.0000	0.0000	0.0000	0
6	−4.60	−4.65	0.0000	0.0000	0.0000	0
7	−4.50	−4.55	0.0000	0.0000	0.0000	0
8	−4.40	−4.45	0.0000	0.0000	0.0000	0
9	−4.30	−4.35	0.0000	0.0000	0.0000	0
10	−4.20	−4.25	0.0000	0.0000	0.0000	0
11	−4.10	−4.15	0.0001	0.0000	0.0000	0
12	−4.00	−4.05	0.0001	0.0000	0.0000	0
13	−3.90	−3.95	0.0002	0.0000	0.0000	0
14	−3.80	−3.85	0.0002	0.0000	0.0001	0
15	−3.70	−3.75	0.0004	0.0000	0.0001	0
16	−3.60	−3.65	0.0005	0.0001	0.0002	0
17	−3.50	−3.55	0.0007	0.0001	0.0002	0
18	−3.40	−3.45	0.0010	0.0001	0.0003	0
19	−3.30	−3.35	0.0015	0.0001	0.0005	0
20	−3.20	−3.25	0.0020	0.0002	0.0007	0
21	−3.10	−3.15	0.0028	0.0003	0.0010	0
22	−3.00	−3.05	0.0038	0.0004	0.0013	0
23	−2.90	−2.95	0.0051	0.0005	0.0019	1
24	−2.80	−2.85	0.0069	0.0007	0.0025	1
25	−2.70	−2.75	0.0091	0.0009	0.0035	1
26	−2.60	−2.65	0.0119	0.0012	0.0046	1
27	−2.50	−2.55	0.0154	0.0015	0.0062	2
28	−2.40	−2.45	0.0198	0.0020	0.0082	2
29	−2.30	−2.35	0.0252	0.0025	0.0107	3

D3 セルには「＝C3*0.1」を入力して，上で求めた確率密度関数の値と幅0.1との積である相対確率を求める。E3 セルには，「＝E2＋D3」を入力し，E2 セルには，あらかじめ，初期値として 0 を入力しておく。これにより，E 列には累積確率が求められることになる。F2 セルには「＝D2*1000」を入力する。これは，D 列で求められた相対確率を1000倍することを意味する。このため，D2 に 0 を入れておく。つまり，1000個のデータがあった場合，各相対確率に相当するデータがいくつあるかを求めることになる。A3 をクリックして F3 までドラッグして，フォーカスした状態を作る。このフォーカスした領域の右下にある黒四角の点，■にマウスの白十字のカーソルを持っていくと，それが黒十字に代わる。その状態で F102 までドラッグする。これによ

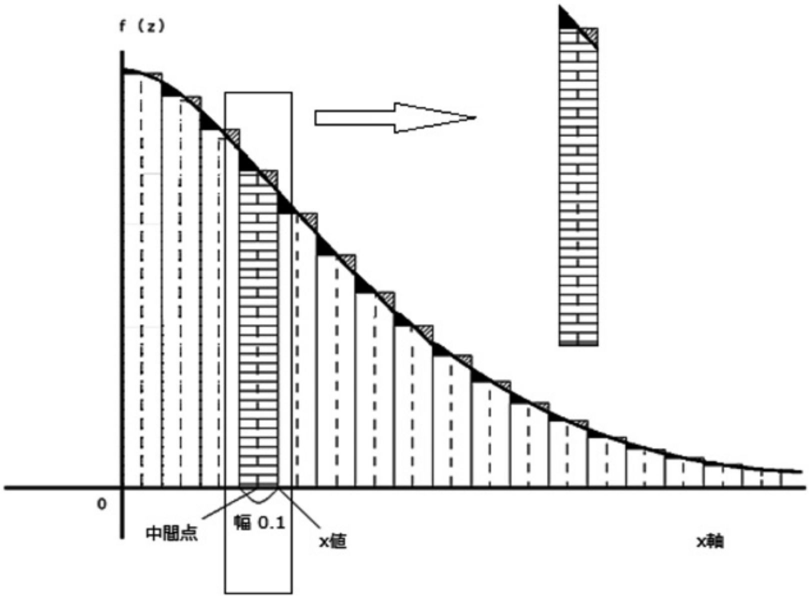

Figure 3. A relative probability of normal distribution.

Table 2. Usage of Canonical normal distribution density function in C3 cell

C3 | =EXP(-B3*B3/2)/SQRT(2*PI())

	A	B	C	D	E	F	G
1	階級	X	確率密度	確率	累積確率	個数	
2	−5.00			0.0000	0.0000	0	
3	−4.90	−4.95	0.0000	0.0000	0.0000	0	
4	−4.80	−4.85	0.0000	0.0000	0.0000	0	
5	−4.70	−4.75	0.0000	0.0000	0.0000	0	
6	−4.60	−4.65	0.0000	0.0000	0.0000	0	
7	−4.50	−4.55	0.0000	0.0000	0.0000	0	
8	−4.40	−4.45	0.0000	0.0000	0.0000	0	
9	−4.30	−4.35	0.0000	0.0000	0.0000	0	
10	−4.20	−4.25	0.0000	0.0000	0.0000	0	
11	−4.10	−4.15	0.0001	0.0000	0.0000	0	
12	−4.00	−4.05	0.0001	0.0000	0.0000	0	
13	−3.90	−3.95	0.0002	0.0000	0.0000	0	
14	−3.80	−3.85	0.0002	0.0000	0.0001	0	
15	−3.70	−3.75	0.0004	0.0000	0.0001	0	
16	−3.60	−3.65	0.0005	0.0001	0.0002	0	

Table 3. Focusing on A3 cell to F3 cell

A3 　 f_x −4.9

	A	B	C	D	E	F
1	階級	×	確率密度	確率	累積確率	個数
2	−5.00			0.0000	0.0000	0
3	−4.90	−4.95	0.0000	0.0000	0.0000	0
4	−4.80	−4.85	0.0000	0.0000	0.0000	0
5	−4.70	−4.75	0.0000	0.0000	0.0000	0
6	−4.60	−4.65	0.0000	0.0000	0.0000	0
7	−4.50	−4.55	0.0000	0.0000	0.0000	0
8	−4.40	−4.45	0.0000	0.0000	0.0000	0
9	−4.30	−4.35	0.0000	0.0000	0.0000	0
10	−4.20	−4.25	0.0000	0.0000	0.0000	0
11	−4.10	−4.15	0.0001	0.0000	0.0000	0
12	−4.00	−4.05	0.0001	0.0000	0.0000	0
13	−3.90	−3.95	0.0002	0.0000	0.0000	0
14	−3.80	−3.85	0.0002	0.0000	0.0001	0
15	−3.70	−3.75	0.0004	0.0000	0.0001	0
16	−3.60	−3.65	0.0005	0.0001	0.0002	0

って，Table 4. で示されているように，A 列に−5.00から5.00までの範囲で，B 列には0.1刻みの中間点，C 列には確率密度，D 列には確率密度，E 列には相対確率，F 列にはサンプル数が1000の場合の相対頻度が求められることになる。

4．正規分布における，一様分布を利用した標本抽出

上述の操作により，1000個のデータを正規分布にしたがって分配した場合の相対度数が得られた。この中から，χ^2 分布の自由度に対応する個数だけサンプリングする必要がある。サンプリングの方法としては，RAND 関数を用いることができる。RAND 関数は，上述の PI( ) 関数同様，引数を必要としない関数で，０以上１以下の数値が得られる。しかし，RAND 関数で得られる数値は，一様分布なので，そのままで用いることができない。そのため，一様分布でも，正規分布に対応したサンプリングができるようにしなければならない。

H 列に１～1000までを割りふる。I 列には，F 列で得られた個数だけ，中間点を入力していく。B23 セルの内容，−2.95で，F23 セルに１が現れる。この内容を I1 セルに入れる。B24 セルの内容，−2.85の個数も１になっているので，I2 セルに−2.85を入れる。同様に，−2.75（B25 セル）を I3 セル，−2.65（B26 セル）を I4 セルに入れる。B27 セルの−2.55は個数が２（F27 セル）なので，I5 セルと I6 セルとに−2.55を入れる。

このように，中間点を表す B 列の値を，同じ行の F 列の値だけくりかえし I 列に入力していく。その結果，I 列には B 列で示された値が F 列の数だけ I 列に書きこまれる。最終的には，I 列に1000個の中間値が並び，これが，確率変数として用いられることになる。これは，近似的に平均 μ が０で，分散 σ が１の分布となる。このようにして，RAND 関数を使った一様分布により正規分布にしたがった標本抽出をすることができるようになる。

Table 4. Normal distribution table : -5.00 to -2.30 (left) and 2.10 to 5.00 (right)

A3　fx　-4.9

	A	B	C	D	E	F
1	階級	×	確率密度	確率	累積確率	個数
2	-5.00			0.0000	0.0000	0
3	-4.90	-4.95	0.0000	0.0000	0.0000	0
4	-4.80	-4.85	0.0000	0.0000	0.0000	0
5	-4.70	-4.75	0.0000	0.0000	0.0000	0
6	-4.60	-4.65	0.0000	0.0000	0.0000	0
7	-4.50	-4.55	0.0000	0.0000	0.0000	0
8	-4.40	-4.45	0.0000	0.0000	0.0000	0
9	-4.30	-4.35	0.0000	0.0000	0.0000	0
10	-4.20	-4.25	0.0000	0.0000	0.0000	0
11	-4.10	-4.15	0.0001	0.0000	0.0000	0
12	-4.00	-4.05	0.0001	0.0000	0.0000	0
13	-3.90	-3.95	0.0002	0.0000	0.0000	0
14	-3.80	-3.85	0.0002	0.0000	0.0001	0
15	-3.70	-3.75	0.0004	0.0000	0.0001	0
16	-3.60	-3.65	0.0005	0.0001	0.0002	0
17	-3.50	-3.55	0.0007	0.0001	0.0002	0
18	-3.40	-3.45	0.0010	0.0001	0.0003	0
19	-3.30	-3.35	0.0015	0.0001	0.0005	0
20	-3.20	-3.25	0.0020	0.0002	0.0007	0
21	-3.10	-3.15	0.0028	0.0003	0.0010	0
22	-3.00	-3.05	0.0038	0.0004	0.0013	0
23	-2.90	-2.95	0.0051	0.0005	0.0019	1
24	-2.80	-2.85	0.0069	0.0007	0.0025	1
25	-2.70	-2.75	0.0091	0.0009	0.0035	1
26	-2.60	-2.65	0.0119	0.0012	0.0046	1
27	-2.50	-2.55	0.0154	0.0015	0.0062	2
28	-2.40	-2.45	0.0198	0.0020	0.0082	2
29	-2.30	-2.35	0.0252	0.0025	0.0107	3

73	2.10	2.05	0.0488	0.0049	0.9822	5
74	2.20	2.15	0.0396	0.0040	0.9861	4
75	2.30	2.25	0.0317	0.0032	0.9893	3
76	2.40	2.35	0.0252	0.0025	0.9918	3
77	2.50	2.45	0.0198	0.0020	0.9938	2
78	2.60	2.55	0.0154	0.0015	0.9954	2
79	2.70	2.65	0.0119	0.0012	0.9965	1
80	2.80	2.75	0.0091	0.0009	0.9975	1
81	2.90	2.85	0.0069	0.0007	0.9981	1
82	3.00	2.95	0.0051	0.0005	0.9987	1
83	3.10	3.05	0.0038	0.0004	0.9990	0
84	3.20	3.15	0.0028	0.0003	0.9993	0
85	3.30	3.25	0.0020	0.0002	0.9995	0
86	3.40	3.35	0.0015	0.0001	0.9997	0
87	3.50	3.45	0.0010	0.0001	0.9998	0
88	3.60	3.55	0.0007	0.0001	0.9998	0
89	3.70	3.65	0.0005	0.0001	0.9999	0
90	3.80	3.75	0.0004	0.0000	0.9999	0
91	3.90	3.85	0.0002	0.0000	1.0000	0
92	4.00	3.95	0.0002	0.0000	1.0000	0
93	4.10	4.05	0.0001	0.0000	1.0000	0
94	4.20	4.15	0.0001	0.0000	1.0000	0
95	4.30	4.25	0.0000	0.0000	1.0000	0
96	4.40	4.35	0.0000	0.0000	1.0000	0
97	4.50	4.45	0.0000	0.0000	1.0000	0
98	4.60	4.55	0.0000	0.0000	1.0000	0
99	4.70	4.65	0.0000	0.0000	1.0000	0
100	4.80	4.75	0.0000	0.0000	1.0000	0
101	4.90	4.85	0.0000	0.0000	1.0000	0
102	5.00	4.95	0.0000	0.0000	1.0000	0
103						1000
104						

5．標本抽出

サンプリング方法は，L列に「＝ROUND(1000*RND( )＋1,0)」を入れる。ROUND関数は，指定した桁数で四捨五入した数値を求める関数で，ROUND(a,b)という形式で用いられる。引数aは，四捨五入すべき数値を表し，引数bは四捨五入したのちの戻り値を何桁で表示するか，つまり四捨五入する位置を指定することができる。bが整数の場合は，小数点以下の桁数，負数の場合は，整数の指定桁に四捨五入する。0の場合は，最も近い整数を返す。これにより，1～1000までの整数が得られる。たとえば，L1セルを見ると乱数により発生した整数が入っている。この整数に対応する中間点の値がM1セルに入れられている。これは，L列の所定のセルで発生した乱数をH列の1～1000の中から検索して，それに対応する中間点を得ることができる。N列には，M列で得られた中間点を正規分布により抽出された確率変数値として利用することができる。

M列のL列に対応する行には，VLOOKUP関数の「＝VLOOKUP(L1,$H$1:$I$1000,2)」を入力する。VLOOKUP関数は，VLOOKUP（検索値，参照範囲，列番号（，検索方法））という形式で，参照範囲の中から検索値と同じものを見つけ，参照範囲の何列目の内容を返すかを決める。検索方法はTRUEかFALSEを選ぶ。省略も可能で，省略した場合はTRUEを指定したのと同じで，検索値の近似値を含めた場合に指定，FALSEを指定した場合は，検索値に完全一致する場合に指定する。

N列にはM列の値を2乗した値を求める計算式が入れられる。たとえば，N1セルには「＝M1*M1」が入れられる。ここで，定義式に戻ってχ^2値を求める。上述のように，平均μが0で，分散σが1の分布であるので，定義式において$\mu=0$，$\sigma=1$を代入すると，$\chi^2=\sum_{i=1}^{n} x_i^2$となり，確率変数$x_i$の2乗和が$\chi^2$値となる。

ただし，$n = 1$ の場合，x_i^2 が χ^2 値となる。これで１つのケースが得られたわけであるが，このような χ^2 値がどのように分布するかを知るためには多くのケースを求める必要がある。ここでは，このようなケースを1000個作ることにし，その分布を見てみることにした。

P 列には，番号として，P2 セルから下に１～1000を入れておく。Q 列には，Q2 から順に，N 列の合計が入力されているセルの行番号が入力される。最初に抽出された＃１のケースでの確率変数の２乗が N2 セルに求められる。２番目に抽出された＃２のケースでの確率変数の２乗が N5 セルに求められる。＃３の合計が N8 セルに，＃４の合計が N11 セルというように３行ごとに表わされる。この順序を反映するように，Q3 セルには，「＝Q2＋3」という数式が入れられる。

自由度が１の場合は，変数の２乗した値がそのまま合計となるので，わざわざ合計欄を作る必要はないのだが，これは，自由度が２以上の場合の形式と揃えるためである。自由度が２以上の場合はこの合計欄に確率変数の２乗の合計が求められる。

R 列には，Q 列に示された合計の行数と文字 N とを組み合わせる CONCATENATE 関数を用いる。CONCATENATE 関数は，形式としては CONCATENATE（文字列１，文字列２，…）という形式で，機能としては文字列１，文字列２，…を結合する役割を果たす。

S 列には N 列にある各ケースの合計を集める。Q 列にある合計の行数とその列名の N を合わせて作られたセルのセル位置名を利用する。S 列には INDIRECT 関数が用いられる。INDIRECT 関数は，形式としては INDIRECT（参照セル位置）という形式で，機能としては参照セル位置の内容を表示する役割を果たす。たとえば，S2 セルには「＝INDIRECT(R2)」が入力されている。これは，参照セル位置 R2 にあるセル位置名 N2 を介して間接的に参照することを意味する。これによって，S2 セルには，N2 セルにある数値が S2 セルに表示されることになる。

6．度数分布表の作成

次に，S 列に並んだ χ^2 値の度数分布表を作成する。ここで，FREQUENCY 関数を利用する。FREQUENCY 関数は次のように使う。１）まず所与のデータの最大値と最小値を含む区間範囲を作成する。U 列に度数分布の階級を表す数値を入れる。ここでは階級幅を２として，32までを U2 から U17 まで入れる。２）次に，区間範囲に隣接する領域に，度数を表示する範囲を設定する。区間範囲が縦列なら，その右側に，区間範囲が横行なら，その下に範囲を設定し，U 列の２～17行目に対応するように，フォーカスした状態で，Figure 4. にあるように，シートとツールバーの間にある *fx* マークをクリックして FREQUENCY を指定する。すると，データ範囲と区間範囲を指定できるようになる。３）指定が終わるとすぐに「OK」ボタンをクリックせずに，Ctrl + Shift を押してから「OK」ボタンをクリックするか，または，Ctrl + Shift + Enter を押す。とすると，区

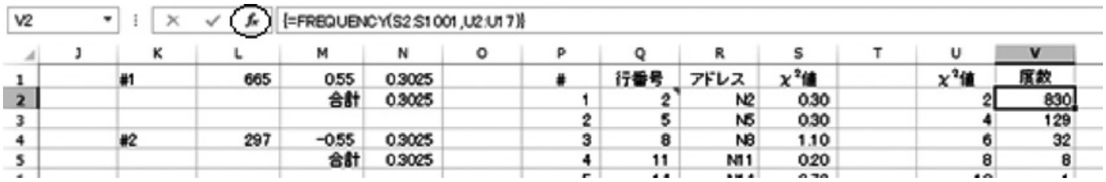

Figure 4. Selection of FREQUENCY function.

間範囲に隣接する領域に，度数分布が表示されることになる。その結果，U2 から U17 までの範囲に {＝FREQUENCY(S2:S1001,U2:U17)} が入る。

以上のような手続きで作られたシートの名前を「df_1」とする。これを利用して「df_2」を作る。「df_1」を右クリックして，現れたポップアップメニューの中から「移動またはコピー」を選択する。そのとき現れたポップアップメニューが Figure 6. に示されている。この中から，下にある「コピーを作成する」の左にあるチェックボックスをクリックしてチェックを入れて，OK ボタンをクリックする。その結果，「df_1(2)」が作成されるので，この名前を変更して「df_2」とする。

df_2 シートでは，自由度が 2 の場合の χ^2 分布を作成する。A 列から K 列まではそのまま利用する。L1 セルから N1 セルまでをコピーし，L2 セルにカーソルを合わせて，右クリックして「コピーしたセルの挿入」を選び，現れたポップアップメニューの「下方向にシフト」を選び，OK ボタンをクリックする。次に，N3 セルにある計算式「＝SUM(N1:N1)」を「＝SUM(N1:N2)」として確率変数となる χ^2 値を 2 つ足し合わせる。区切りのために，K3 セルに空欄を入れる。これで，K1 セルから N4 セルまでの範囲をフォーカスした状態で，下に＃1000ケースになるまでドラッグする。

Q2 セルに＃ 1 ケースの合計欄の行数である 3 を入れる。Q3 には「＝Q2＋4」を入れる。この＋4 は各ケースの合計欄の間隔行数を意味する。この後，P3 セルから Q3 セルまでをフォーカスして，これを1001行目までドラッグする。これにより，df_1 と同様，R 列には確率変数の χ^2 値のセル位置が，S 列には確率変数そのものが表示される。

df_3 シートも，df_2 シートの場合と同様に，df_2 シートをコピーしてできた df_2(2) シートの名前を変更して df_3 シートとする。df_3 シートでは，自由度が 3 の場合の χ^2 分布を作成する。A 列から K 列まではそのまま利用する。空欄を含めた K2 セルから N2 セルまでコピーして，K3 にカーソルを合わせて，右クリックして「コピーしたセルの挿入」を選び，現れたポップアップメニューの「下方向にシフト」を選び，OK ボタンをクリックする。さらに，N4 セルにある計算式「＝SUM(N1:N2)」を「＝SUM(N1:N3)」として，確率変数とな

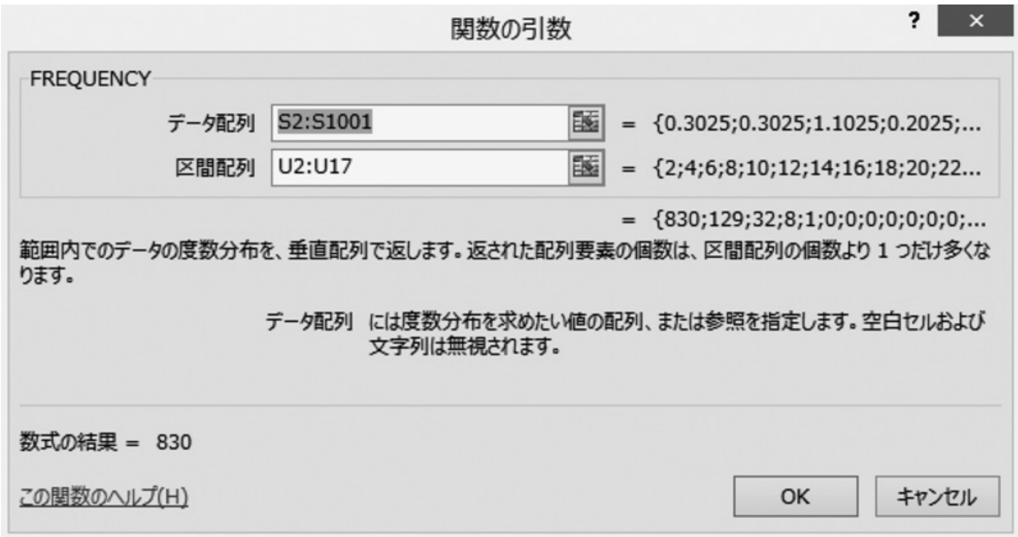

Figure 5. Specification of FREQUENCY function.

Figure 6. Duplication of a sheet.

る χ^2 値を3つ足し合わせる。K1 セルから N5 セルまでの範囲をフォーカスした状態で，下に＃1000ケースになるまでドラッグする。

Q2 セルに＃１ケースの合計欄の行数である４を入れる。Q3 には「＝Q2＋5」を入れる。この ＋5 は各ケースの合計欄の間隔行数を意味する。P3 セルから Q3 セルまでをフォーカスして，これを1001行目までドラッグする。これにより，df_1 と同様，R 列には確率変数の χ^2 値のセル位置が，S 列には確率変数そのものが表示される。

以下同様にして，シートをコピーして，名前を変更し，各シートの Q2 セルに＃１ケースの合計欄の行数を入れる。この値は Table 4. の Q2 の行にシート別に示されている。χ^2 値を求める確率変数の数を Table 4. の d の行に示してあるだけ増加させる。こうして得られた χ^2 値の度数分布表を作成する。

以上に示したように，正規分布にしたがう確率変数の中から自由度の数だけ確率変数を抽出して，それらを２乗したものを足し合わせて χ^2 値を算出する。そして，さらにその χ^2 値の分布を求めることによって度数

Table 4. Initial row number (Q2) and increment (Q3) in Q column of each sheet.

	df_1	df_2	df_3	df_4	df_5	df_6	df_7	df_8	df_9	df_10
Q2	2	3	4	5	6	7	8	9	10	11
Q3	5	7	9	11	13	15	17	19	21	23
d	＋3	＋4	＋5	＋6	＋7	＋8	＋9	＋10	＋11	＋12

注） Q2 行は＃１ケースの合計欄の行数，Q3 行は＃２ケースの合計欄の行数を示す。
D 行は Q3 セルに入力される計算式「＝Q2＋d」における d の値を示す。

Figure 7. Tags of each sheet.

分布表が得られ，それをもとに分布図が描けるようになる。

7．χ^2 分布図の作成

このようにして得られた度数分布表をもとにして，グラフを作成する。χ^2 分布の分布図を表示するためのシート Figure を作成する。先に得られた χ^2 値の確率変数の度数分布表から各自由度の階級別の頻度を表にする。

C 列の C2 セルから C17 セルに，df_1 シートの V2 セルから V17 セルを表示する。このため，Figure シートの C2 セルには「＝df_1! V17」という計算式を入れる。別のシートのセルにある内容を表示する場合には「＝シート名！セル位置」という形式で入力する。

集められた度数分布は Table 5. に示されている。これは，縦列で表示されたものであるが，これを図示するためには，度数分布を横行に表示することになる。そのために，作成されたのが Table 6. である。Table 6. には，たとえば Table 5. の C2 ～ C17 セルに相当する範囲のセル位置が C38 ～ R47 セルに示されている。その内容を Table 7. で，前述の INDIRECT 関数を用いて表示する。これを範囲指定して，ツールバーの「挿入」タブからグラフを選択して折れ線グラフを指定すれば，Figure 8 のような分布図が表示される。

以上のように，χ^2 値の定義から始まり，正規分布の確率密度関数を取り上げ，正規分布での標本抽出法を述べて，その標本抽出法によって得られたデータのもつ度数分布に対して，度数分布表の作成し，それに基づい

Table 5. An example of data table of Chi-square distribution curves in Figure-sheet.

Q17

	A	B	C	D	E	F	G	H	I	J	K	L
1			df=1	df=2	df=3	df=4	df=5	df=6	df=7	df=8	df=9	df=10
2		2	803	620	429	267	144	67	44	30	11	5
3		4	130	227	322	333	284	222	168	123	80	56
4		6	40	89	139	211	248	264	240	204	159	133
5		8	20	44	74	92	156	195	205	201	219	171
6		10	7	17	27	60	83	126	143	166	170	168
7		12	0	3	3	25	54	67	103	118	126	166
8		14	0	0	5	9	14	35	46	86	130	129
9		16	0	0	1	1	11	12	24	35	47	72
10		18	0	0	0	2	4	7	17	20	26	53
11		20	0	0	0	0	2	3	5	8	19	19
12		22	0	0	0	0	0	2	2	5	7	17
13		24	0	0	0	0	0	0	3	3	3	3
14		26	0	0	0	0	0	0	0	1	2	6
15		28	0	0	0	0	0	0	0	0	0	0
16		30	0	0	0	0	0	0	0	0	0	1
17		32	0	0	0	0	0	0	0	0	1	1
18												

Table 6. An example data table of cell positions of Chi-square distribution in Figure-sheet.

C2　=df_1!V2

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
36																			
37		χ square	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	
38		df=1	825	119	41	11	4	0	0	0	0	0	0	0	0	0	0	0	
39		df=2	631	225	96	34	12	1	0	1	0	0	0	0	0	0	0	0	
40		df=3	430	281	190	63	22	12	1	1	0	0	0	0	0	0	0	0	
41		df=4	262	325	183	132	58	25	13	2	0	0	0	0	0	0	0	0	
42		df=5	153	295	230	146	105	50	15	3	3	0	0	0	0	0	0	0	
43		df=6	100	229	252	188	106	76	26	13	6	4	0	0	0	0	0	0	
44		df=7	40	172	229	195	143	102	58	36	18	5	2	0	0	0	0	0	
45		df=8	23	123	202	215	160	118	79	42	28	6	4	0	0	0	0	0	
46		df=9	7	74	185	210	202	147	66	52	35	15	5	1	1	0	0	0	
47		df=10	4	46	135	196	181	154	109	75	52	26	12	8	0	0	1	1	
48																			

Table 7. Converted data table of Chi-square distribution curves in Figure-sheet.

χ square	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32
df=1	825	119	41	11	4	0	0	0	0	0	0	0	0	0	0	0
df=2	631	225	96	34	12	1	0	1	0	0	0	0	0	0	0	0
df=3	430	281	190	63	22	12	1	1	0	0	0	0	0	0	0	0
df=4	262	325	183	132	58	25	13	2	0	0	0	0	0	0	0	0
df=5	153	295	230	146	105	50	15	3	3	0	0	0	0	0	0	0
df=6	100	229	252	188	106	76	26	13	6	4	0	0	0	0	0	0
df=7	40	172	229	195	143	102	58	36	18	5	2	0	0	0	0	0
df=8	23	123	202	215	160	118	79	42	28	6	4	0	0	0	0	0
df=9	7	74	185	210	202	147	66	52	35	15	5	1	1	0	0	0
df=10	4	46	135	196	181	154	109	75	52	26	12	8	0	0	1	1

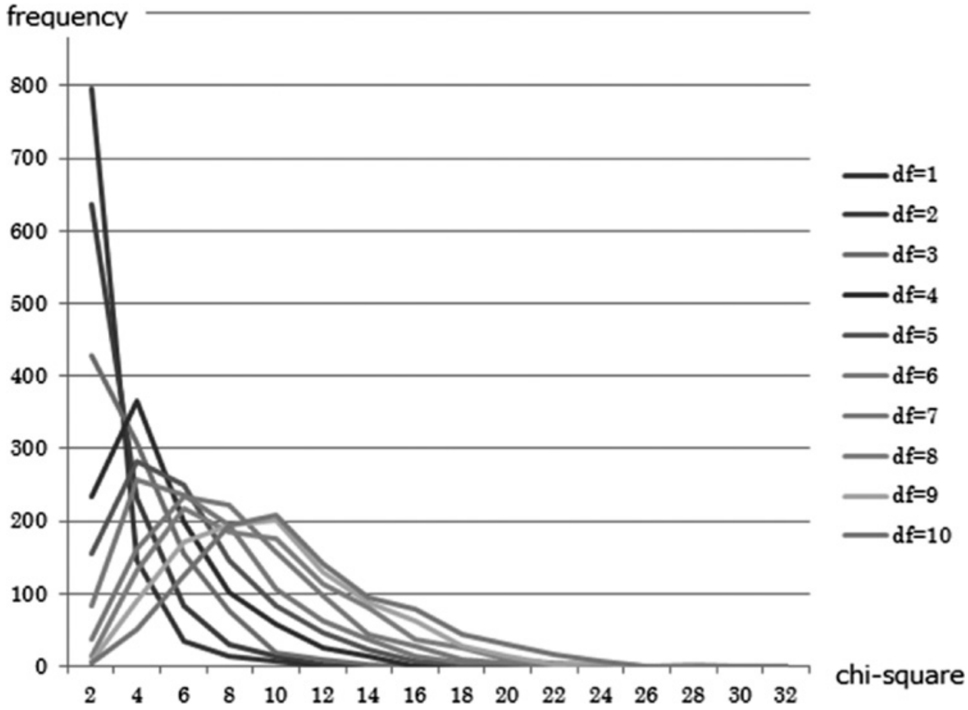

Figure 8. An example of chi-square distribution curves (df=1~10).

て χ^2 分布図の作成をすることができた。このようにして，χ^2 分布の定義から，確率分布図を導き出す過程をEXCEL によるシミュレーションを使って具体的な事象として擬似的に体験することによって直観的な理解を深めることができるようになる。

注

1） OS は Windows8.1 で，ソフトは Microsoft Excel2013 を用い，CPU は Intel Corei7 を用いた。

2） 正規分布表の導出については，門田（2013）を参照のこと。

参考文献

日花弘子　2011 「仕事に役立つ Excel 統計解析　第３版」ソフトバンククリエイティブ

IT フロンティア　2003 「Visual Basic.NET 逆引き大全500の極意」秀和システム

岩原信九郎　1970 「教育と心理のための推計学」日本文化科学社

金城俊哉　2005 「Visual Basic パーフェクトマスター」秀和システム

門田幸太郎　2013 「Excelによるシミュレーションを用いた正規分布表の詳細化とVisual Basicによる累積確率の検索方法」立命館産業社会論集 第48巻 第4号 pp. 123-134
守谷栄一　1987 「詳解演習数理統計」日本理工出版会
村上雅人　2002 「なるほど統計学」海鳴社
武藤真介　1995 「統計解析ハンドブック」朝倉書店
成富慶子　2007 「EXCEL関数辞典」秀和システム
Norma Gilbert　1981 “Statistics second edition” Saunders college publishing
芝祐順・渡部洋・石塚智一　1984 「統計用語辞典」新曜社
高木貞治　1983 「解析概論」岩波書店
竹村彰通　2000 「統計 第2版」共立出版
常見美保　2007 「EXCELVBA辞典」秀和システム
山本昌弘・重定恕彦　2004 「例題でわかるVisual Basic.NET」東京電機大学出版局
和田秀三　1990 「基本演習確率統計」サイエンス社

Simulation Using Excel for an Understanding of Chi-Square Distribution :
The Relationship between the Definition of Chi-Square Valuable Function and the Distribution Curves

MONDEN Kotaro [i]

Abstract : The purpose of this paper is to assist beginners in gaining an intuitive understanding of the process of making a chi-square distribution diagram from the definition of the chi-square distribution using EXCEL simulations.

First, an overview of the usage and logic of the chi-square test is provided, followed by the definition of chi-square value. Then the probability density function of the normal distribution leads a method of sampling for normal distribution using the RAND function in EXCEL. For the frequency distribution obtained as a result of sampling, applying the FREQUENCY function to a frequency table creates distribution diagrams based on it.

Keywords : Canonical Normal Distribution, Chi-Square Distribution, Chi-Square Function, Chi-Square Test, Excel Function, FREQUENCY Function, Probability Density Function, Simulation by Excel, VLOOKUP Function

i Professor, Faculty of Social Sciences, Ritsumeikan University