

9. 検定

テキスト分析やアンケート調査によって求められた和や平均などの数値が「統計的に有意」であるかどうかを調べる方法として様々な検定法が使われます。それらは「確率」の章で扱った確率分布の性格に基づいています。確率には「まったくありえない」確率(0%)から、「絶対にそうなる」という確率(100%)までありえますが、たとえば数値 A と数値 B の間に「差がまったくない(0%)」または「差が 100%ある」と言い切ることは困難です。そこで「差がない、とは言えない」(「言い難い」という消極的な判断をします。つまり「A と B には差がない」という仮説からスタートするのは。この仮説は「帰無仮説」(H_0)とよばれます。無に帰したい(棄却したい)仮説なのでこのようによばれます。この逆の「差がある」という仮説は「対立仮説」(H_1)とよばれます¹。

帰無仮説(H_0): 「A と B には差がない」 (偶然でも生じる)

対立仮説(H_1): 「A と B には差がある」 (偶然では生じにくい)

和や平均などの統計量が偶然に生じること(帰無仮説)ではないことであれば、帰無仮説(H_0)が棄却(reject)され、対立仮説(H_1)が採用(accept)されます。和や平均などの統計量が偶然には生じない領域は、その確率分布のグラフの中の「棄却域」(critical region)とよばれ、ふつう全体の 5%または 1%という小さな領域が設定されます。その領域の大きさを決める基準は「有意水準」(significance level)とよばれます。これは対立仮説を支持するケースの割合を示し、パーセントや小数で示され、一般に 5%と 1%が用いられます。

たとえば「5%の有意水準で帰無仮説 H_0 が棄却される」という結論は、 H_0 の成り立つ確率が 5%以下 (H_1 が成り立つ確率が 95%以上)ということになります。逆に言えば帰無仮説を 5%の有意水準で棄却すると、本当は帰無仮説が正しいのに、誤って対立仮説を「楽観的に」採用してしまうこととなります(「第 1 種の誤り」(type I error))²。それを避けるには有意水準をさらに厳しくして 1%や 0.1%にすることも考えられるでしょう。そうすると、今度は本当は対立仮説が正しいのに(差がある)、基準を厳しくしたために帰無仮説を「悲観的に」採用してしまう(差がない)こととなります(「第 2 種の誤り」(type II error))。

そこで、たとえば有意水準(α)を 5%にして結論を導くための絶対的な基

¹ ふつう対立仮説の有意性を問題にするので、これが研究の作業仮説になります。一方、たとえば頻度の分布が従来の資料・説による分布と同じであるという作業仮説(帰無仮説)が成立することを確かめる研究もあります。そのときは帰無仮説の有意水準を厳しく(小さく)する必要があります(一般の有意水準を 95%や 99%に設定します)。

² そのため「有意水準」は「危険率」(risk ratio)ともよばれます。

準とすると、たとえば統計量の確率(p)が 4.99%であれば、帰無仮説を棄却し、5%であれば帰無仮説を採択する、ということになります。実際に 5%と 4.99%の違いは本質的ではないはずで、よって有意水準は一応の「めやす」としての基準であって、むしろ実際の統計量の確率(p)の大きさ(小ささ)と帰無仮説の採択域(「信頼区間」 confidence interval)の幅の様子を連続的に観察すべきでしょう。

以下では簡単な「二項検定」からはじめ、少しずつ段階的に各種の検定法を見ていきます。それぞれの検定法を理論的に「理解」することとは別に、実際に実験して「納得」しながらその意味を確かめます。

9.1. 二項検定

9.1.1. 二項分布

一定の確率で起こる現象が、n 回の中で、x 回起きたことが統計的に有意であるかどうかを検定する方法として、簡単な「二項検定」(Binary test)が使われます。たとえばコインを 10 回投げて 8 回「表」が現れたとします。このことがどのような確率になるかを二項確率の関数式 Bin で求めると(→「確率」)³

$$\text{Bin}(x, n, p) = \text{Bin}(8, 10, 0.5) = .0439$$

ここで x はコインの表が現れた回数、n は全回数、0.5 は事前確率を示します。コインの表が出る確率が 0.5 であることは明らかです。よって、確率が 0.5 の事象(コインの表が出ること)が、全 10 回の中で 8 回起こる、という確率(事後確率)は 4.4%ぐらいだと言えます。よってかなり稀な現象だと思われそうですが、さらに珍しいケースであると思われる 9, 10 回の「表」が現れるケースの確率も足し上げていきます。これは「累積確率」(Cumulative probability: CP)とよばれます⁴。

$$\text{CP}(x, n, p) = \sum [i:x, n] \text{Bin}(i, n, p)$$

$$\text{CP}(24, 39, 0.5) = .0547$$

³ Excel では BINOMDIST(x, n, p, 0)という関数を使います。この関数は次の二項分布確率の式で求める値と同じになります(→「確率」「二項分布」)。

$$\text{Binom}(x, n, p) = {}_n\text{C}_x p^x (1-p)^{n-x} = n! / [r! (n-r)!] p^x (1-p)^{n-x}$$

ここで n はデータ数、x は生起数、p は生起する確率を示します。

⁴ Excel では累積確率を BINOMDIST(x, n, p, 1)という関数で求めます。よって、有意水準・危険率 Sig.は

$$\text{Sig.}(x, n, p) = 1 - \text{BINOMDIST}(x-1, n, p, 1)$$

を使います。BINOMDIST(x-1, n, p, 1)で生起数 x に満たないすべての場合を足し上げ、それを全体の確率(=1)から引けば、それが生起数 x 以上のすべての場合の確率の和になります。ただし、x=0 のときは計算ができないので、はじめに返す値を 1 としておきます。

上式の右式は二項確率(bin)の第1引数*i*を*x*から*n*まで1つずつ増やしながら足し上げていくことを示します。ここで足し上げるときの*i*の初期値が*x=8*であることに注意してください。この*x*が8, 9, 10まで続きます。それらの引数で毎回計算された二項確率の合計は.0547になります。この数値（当該の確率.046より上・右側の確率の和）は「上側確率」(Upper probability: UP)とよばれます。

$$UP(x,n,p) = CM(x, n, p) = \sum [i:x,n] Bin(i, n, p)$$

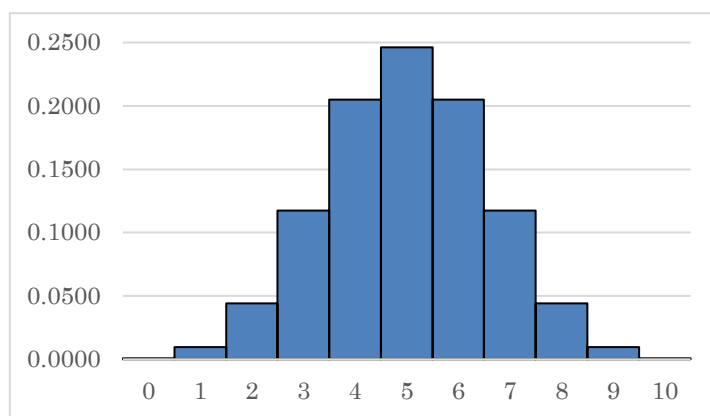
$$UP(24,39,1/2) = \sum [i:24,39] Bin(i, 39, 1/2) = .054$$

つまり、このコインは「表」が出る割合が「裏」が出る割合と差がない、という帰無仮説を棄却すると5.47% (.054)の確率で誤りがある、という危険をおかすことになります。「表」が現れるケースは全体の5%以上もあるからです。検定ではふつう有意水準を5%（または1%）に定めます。よって、このコインは「表」と「裏」が出る割合の差がない、という帰無仮説を5%の有意水準で棄却することができません。

次の表は、*x*の値を0, 1, ..., 10まで上げながら、有意水準(Sig.)を計算したものです。これを見ると危険確率(Risk)が5%を切るのは*x=9*であることがわかります⁵。

x	n	e.p	Bin.p	Lower p.	Upper p.
0	10	0.500	0.0010	0.0010	1.0000
1	10	0.500	0.0098	0.0107	0.9990
2	10	0.500	0.0439	0.0547	0.9893
3	10	0.500	0.1172	0.1719	0.9453
4	10	0.500	0.2051	0.3770	0.8281
5	10	0.500	0.2461	0.6230	0.6230
6	10	0.500	0.2051	0.8281	0.3770
7	10	0.500	0.1172	0.9453	0.1719
8	10	0.500	0.0439	0.9893	0.0547
9	10	0.500	0.0098	0.9990	0.0107
10	10	0.500	0.0010	1.0000	0.0010

⁵ Bin = BINOMDIST(x, n, p, 0)
 Lower p. =BINOMDIST(x, n, p,1)
 Upper p. = 1 - BINOMDIST(x-1, n, p, 1)



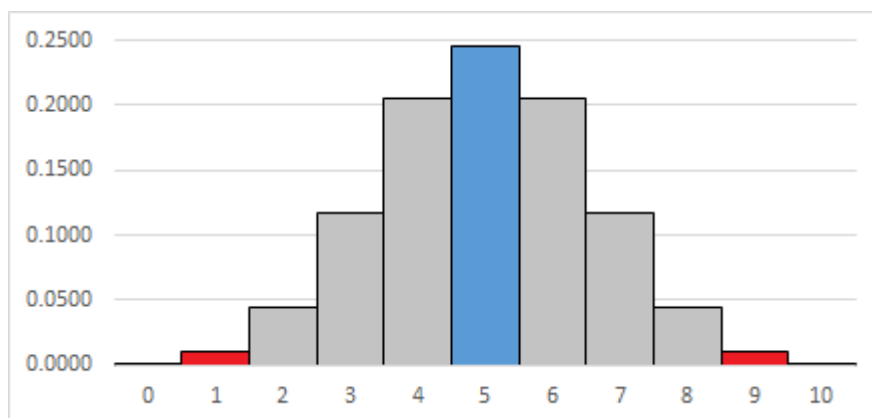
私たちは言語データの頻度分布を観察するときには検定をするばかりでなく、一般にその頻度がどの程度信頼できるのかを知る必要があります。たとえば、24, 15 という 2 つの数値を比べて、24の方が有意に大きい、ということ、確率的に判定することができるでしょうか？これを二項検定を使って判定します。15 に比べて 24 は有意に大きいかどうか、ということ、を、39 (=24+15) の中で 23 は有意に大きいか、という問題に置き換え、 $x = 24$, $n = 39$, $p = 1/2$ における上側確率を求めると

$$\text{Upper } p = 1 - \text{BINOMDIST}(23, 39, 1/2, 1) = .054$$

よって、24 は 15 と比べて有意に大きいとは言えない、ということになります。

●二項分布の最大項

本文の表を見ると $\text{Bin}(5, 10, .500)$ が最大の確率（頂点）を示しています ($\text{Bin}.p = .2461$)。このときの $\text{Bin}.p$ の値(.2461)を「最大確率」とよび、 $x = 5$ を「最大項」とよびましょう。次のグラフの青の棒が最大確率を示します。



このとき、最大項が 5 になることは、期待確率が 0.5 のときに試行回数 n が 10 であれば、平均は $10 * 0.5 = 5$ となるので当然だと思われま。たしかに、多くの場合、最大項は平均値の位置と同じになりますが、平均値が小数点を持つとき、最大項は整数なので、平均値を丸める方法（四捨五

入・切り上げ・切り捨て) を考えなければなりません。また、後で見るように、平均値が必ずしも最大項を決定するとは限りません。

そこで、数理的に最大項を求める方法を探ります。はじめに、連続する二項分布確率 $P(x)$ と $P(x+1)$ の比 $P(x+1) / P(x)$ を求めます。

$$P(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$P(x+1) = \binom{n}{x+1} p^{x+1} (1-p)^{n-x-1}$$

$$P(x+1) / P(x) = \frac{\binom{n}{x+1} p^{x+1} (1-p)^{n-x-1}}{\binom{n}{x} p^x (1-p)^{n-x}}$$

上式の右辺を次の (a) と (b) に分解します。

$$P(x+1) / P(x) = (a) * (b)$$

$$(a) = \frac{\binom{n}{x+1}}{\binom{n}{x}}, (b) = \frac{p^{x+1} (1-p)^{n-x-1}}{p^x (1-p)^{n-x}}$$

$$\begin{aligned} (a) &= \frac{\binom{n}{x+1}}{\binom{n}{x}} \\ &= \frac{\{n! / [(x+1)! (n-x-1)!]\}}{\{n! / [x! (n-x)!]\}} \\ &= \frac{\{n! [x! (n-x)!] / n! [(x+1)! (n-x-1)!]\}}{\{n! [(x+1)! (n-x-1)!] / n! [x! (n-x)!]\}} \\ &= \frac{[x! (n-x)!]}{[(x+1)! (n-x-1)!]} \\ &= \frac{[x! (n-x) (n-x-1)!]}{[(x+1) x! (n-x-1)!]} \\ &= (n-x) / (x+1) \end{aligned}$$

$$\begin{aligned} (b) &= \frac{p^{x+1} (1-p)^{n-x-1}}{p^x (1-p)^{n-x}} \\ &= \frac{[p * p^x * (1-p)^{n-x-1}]}{[p^x * (1-p) * (1-p)^{n-x-1}]} \\ &= p / (1-p) \end{aligned}$$

よって、

$$\begin{aligned} P(x+1) / P(x) &= (a) * (b) \\ &= [(n-x) / (x+1)] * [p / (1-p)] \\ &= [(n-x) p] / [(x+1) (1-p)] \end{aligned}$$

二項確率分布のグラフを見ると、最大確率までが順調に上昇し、最大確率を超えると順調に下降していることがわかります。よって、最大確率の前は $P(x+1) > P(x)$ の関係が成立し、最大確率の後には $P(x+1) < P(x)$ の関係が成立します。はじめに $P(x+1) > P(x)$ の関係を見ます。

$$\begin{aligned} P(x+1) > P(x) &\Leftrightarrow P(x+1) / P(x) > 1 && \leftarrow P(x) > 0 \\ &\Leftrightarrow [(n-x) p] / [(x+1) (1-p)] > 1 \\ &\Leftrightarrow (n-x) p > (x+1) (1-p) && \leftarrow (x+1) (1-p) > 0 \\ &\Leftrightarrow np - xp > x - xp + 1 - p \\ &\Leftrightarrow np > x + 1 - p \\ &\Leftrightarrow x < np + p - 1 \end{aligned}$$

同様にして，最大確率の後は，

$$\begin{aligned} P(x+1) < P(x) &\Leftrightarrow P(x+1) / (pi) < 1 \\ &\Leftrightarrow x > np + p - 1 \end{aligned}$$

よって，最大項(x)は $np + p - 1$ になります。たとえばコインを 10 回投げて，最大の確率となるのは， $10 * 0.5 + 0.5 - 1 = 4.5$ の位置です。ところが二項分布確率グラフの横軸はすべて整数になるので，4.5 の位置はありません。ここで次の関係に注目しましょう。

$$\begin{aligned} P(x+1) > P(x) &\Leftrightarrow x < np + p - 1 = 4.5 \\ P(x+1) < P(x) &\Leftrightarrow x > np + p - 1 = 4.5 \end{aligned}$$

上の式は「 $x < 4.5$ ならば確率が上昇し， $x > 4.5$ ならば下降する」ということを示します。よって $x = 4$ ならば 4.5 以下なので確率が上昇して $P(4) < P(5)$ となり， $x = 5$ ならば 4.5 以上なので，確率は下降して， $P(5) > P(6)$ になります。つまり，最大項(m)は $x = np + p - 1$ を切り上げた整数です。

$$m = \text{RoundUp}(x) = \text{RoundUp}(np + p - 1)$$

全体の確率は次のような分布を示し，最大確率は $P(5)$ となります。

$$P(0) < P(1) < \dots < P(4) < \mathbf{P(5)} > P(6), \dots, P(10)$$

さて， x が整数値になる場合があります。たとえば， $n = 13, p = 1/2$ のときは，

$$x = np + p - 1 = 13 * 1/2 + 1/2 - 1 = 6$$

このとき先の式にあてはめれば，「 $x < 6$ ならば確率が上昇し， $x > 6$ ならば下降する」ということになるので， $x = 0, 1, \dots, 5$ と $x = 7, 8, \dots, 13$ については，それぞれの確率の向きは上昇と下降になります。そして $x = 6$ のときは，上昇も下降もしないで， $P(x) = P(x+1)$ ，つまり $P(6) = P(7)$ になります。

$$P(0) < P(1) < \dots < P(5) < \mathbf{P(6) = P(7)} > P(8), \dots, P(13)$$

このことを実際に計算して確かめましょう。

x	n	p	Bin	Lower p.	Upper p.
0	13	0.5	0.0001	0.0001	1.0000
1	13	0.5	0.0016	0.0017	0.9999
2	13	0.5	0.0095	0.0112	0.9983
3	13	0.5	0.0349	0.0461	0.9888

4	13	0.5	0.0873	0.1334	0.9539
5	13	0.5	0.1571	0.2905	0.8666
6	13	0.5	0.2095	0.5000	0.7095
7	13	0.5	0.2095	0.7095	0.5000
8	13	0.5	0.1571	0.8666	0.2905
9	13	0.5	0.0873	0.9539	0.1334
10	13	0.5	0.0349	0.9888	0.0461
11	13	0.5	0.0095	0.9983	0.0112
12	13	0.5	0.0016	0.9999	0.0017
13	13	0.5	0.0001	1.0000	0.0001

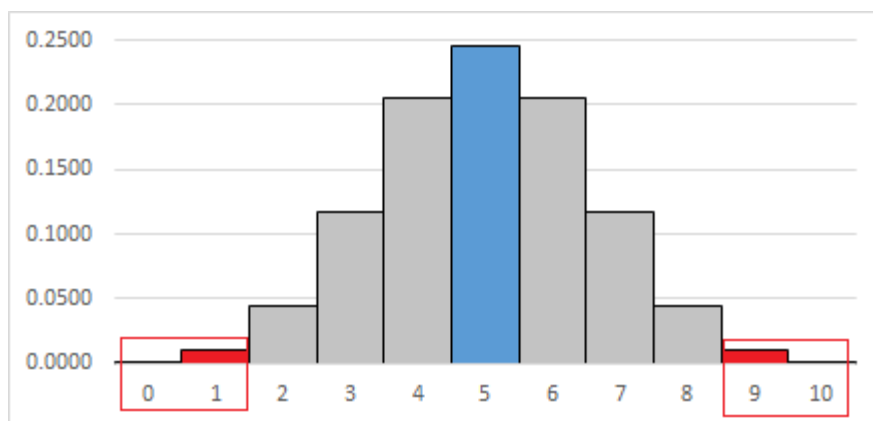
最大項を示す $x = np + p - 1$ の中の np は平均値を示し、 p は確率なので、 $(0, 1)$ の範囲です。そこで、 x は、ほとんど平均値に近いことになります。しかし、たとえば、 $n = 95$ 、 $p = 0.05$ のとき、 $x = 3.8$ になります。このとき、 $np = 4.75$ なので平均値が最大項を示していません。このように p が小さな確率であるときに、平均値以下の位置で最大確率になります。

参照：小寺平治(2002)『ゼロから学ぶ統計解析』講談社 (p.77)

● 帰無仮説の肯定的検定

一般に帰無仮説は棄却されるか、棄却されないかのどちらかであり、棄却されないときでも、帰無仮説が肯定されることはありません。よって、帰無仮説の正否は常に不明、ということになります。そこで、一定の危険率を承知したうえで肯定する方法を探ります。先に使った表とグラフを再掲します。ここではコインを 10 回投げて表が出た回数を x とします。

x	n	e.p	Bin.p	Lower p.	Upper p.	Central p.
0	10	0.500	0.0010	0.0010	1.0000	1.0000
1	10	0.500	0.0098	0.0107	0.9990	0.9980
2	10	0.500	0.0439	0.0547	0.9893	0.9785
3	10	0.500	0.1172	0.1719	0.9453	0.8906
4	10	0.500	0.2051	0.3770	0.8281	0.6563
5	10	0.500	0.2461	0.6230	0.6230	0.2461
6	10	0.500	0.2051	0.8281	0.3770	0.6563
7	10	0.500	0.1172	0.9453	0.1719	0.8906
8	10	0.500	0.0439	0.9893	0.0547	0.9785
9	10	0.500	0.0098	0.9990	0.0107	0.9980
10	10	0.500	0.0010	1.0000	0.0010	1.0000



たとえば有意水準 5%として、 $x = 8$ ならば、その上側確率 $\text{Upper } p.(8, 10, 1/2) = 0.0549$ は 5%以下ではないので、帰無仮説 (H_0 : 表が出る確率が $1/2$ の正しいコインである) を棄却できません。このようなとき、一般に帰無仮説が棄却されない、と言っても、帰無仮説を肯定することはなく、帰無仮説に対する判断は保留されています。つまり、正しいコインであるかどうかについては判断できない、ということです。小寺(2002:140)は次のように述べています。

帰無仮説 H_0 が棄却されないとき、“採択される”といいますが、これは、けっして、帰無仮説が正しいと認めたわけではありません。帰無仮説が棄却されないときは、何の結論も下せないということです。

たしかに、 $x = 8$ ならば、5%以上の危険性があることに有意性を認めることはできませんし、また逆に帰無仮説を肯定することもできません。そのようなことは偶然でも $\text{Lower } p.(8, 10, 1/2) = .9893 = 98.93\%$ の確率で起こることだからです。

一方、確率最大項 $x = 5$ のときは、どうでしょうか？そのようなことが偶然で起こるのは $\text{Bin. } p(5, 10, 1/2) = .2461$ なので 24.61% の確率となります。この確率は、帰無仮説を肯定したときに、それが誤っている確率(危険率)を示します。24.61% という危険率はかなり高いのですが(そのようなことは偶然でも 24.61% の確率で起きる)、それでもこの危険率を認めたらうで、ということならば帰無仮説を肯定することができます。 x が確率最大項に近いほど、そして n が大きくなるほど、帰無仮説採択の危険率は小さくなるはずですが、つまり、あまりに帰無仮説どおりである、ということと比較的小さな危険率で認めることができる、ということです。

ここで、帰無仮説危険率(null hypothesis risk: NHR)を次のように計算します。

$$\text{NHR}(5) = \text{Central } p.(5) = .2461$$

$$\text{NHR}(6) = \text{Central } p.(6) = .2461 + .2051 * 2 = .6563$$

$$\text{NHR}(7) = \text{Central } p.(7) = .2461 + .2051 * 2 + .1171 * 2 = .8906$$

とします。上側も同様です。

$$\text{NHR}(4) = \text{Central p. (4)} = .2461 + .2051 * 2 = .6563$$

$$\text{NHR}(3) = \text{Central p. (3)} = .2461 + .2051 * 2 + .1171 * 2 = .5684$$

とします。これらの数値は当該の確率を超える確率をすべて足し合わせた数値で、最大項を中心として偶然でも起こる確率の和になります。たとえば、NHR(6)の確率は.2051なのですが、その確率と同じ確率をもつ事象、または、それを超える確率をもつ事象が起こる確率を示します(x=4, 5)。帰無仮説危険率は、その事象の確率以上の確率をもつ帰無仮説を肯定したとき、それが過誤であることの危険率を示します。

これらの危険率の有意水準(b)の目安を 50%または 25%にします。50%以上ならば半数以上間違えている可能性があるので肯定の余地はありません。逆に 50%以下ならば正しいと判断する余地が残されています。基準をより厳しくするときには a = 25%が目安です。上のグラフで、x = 0, 1 は下側対立仮説採択域(a = 5%: グラフの赤の左の領域), 9, 10 は上側対立仮説採択域になります(a = 5%: グラフの赤の右の領域), x = 5 を帰無仮説採択域(グラフの青の領域)とし、x = 2, 3, 4, 6, 7, 8 は判断保留域(グラフのグレーの左右の領域)とします。

次は、n = 100 としたときの最大項の周りの領域です(x=41, 42, ..., 59)。このとき、帰無仮説採択域は x = 49, 50, 51 (b = 25%), または x = 48, 49, 50, 51, 42 (b = 50%)になります。この領域の外側は両側(x=47, 53)に中央確率(Central p.) = .5159 があるので、50%以上になってしまい、帰無仮説採択域を脱して、判定保留域に入ります。

x	n	e.p	Bin.p.	Lower p.	Upper p.	Central p.
41	100	0.5	0.0159	0.0443	0.9716	0.9431
42	100	0.5	0.0223	0.0666	0.9557	0.9114
43	100	0.5	0.0301	0.0967	0.9334	0.8668
44	100	0.5	0.0390	0.1356	0.9033	0.8067
45	100	0.5	0.0485	0.1841	0.8644	0.7287
46	100	0.5	0.0580	0.2421	0.8159	0.6318
47	100	0.5	0.0666	0.3086	0.7579	0.5159
48	100	0.5	0.0735	0.3822	0.6914	0.3827
49	100	0.5	0.0780	0.4602	0.6178	0.2356
50	100	0.5	0.0796	0.5398	0.5398	0.0796
51	100	0.5	0.0780	0.6178	0.4602	0.2356
52	100	0.5	0.0735	0.6914	0.3822	0.3827
53	100	0.5	0.0666	0.7579	0.3086	0.5159
54	100	0.5	0.0580	0.8159	0.2421	0.6318

55	100	0.5	0.0485	0.8644	0.1841	0.7287
56	100	0.5	0.0390	0.9033	0.1356	0.8067
57	100	0.5	0.0301	0.9334	0.0967	0.8668
58	100	0.5	0.0223	0.9557	0.0666	0.9114
59	100	0.5	0.0159	0.9716	0.0443	0.9431

これまでは期待確率(e.p.)を 0.5 としてきました。そのため二項確率分布は左右対称になりますが，期待確率(e.p.)を，たとえば 0.2 にすると，その確率最大項は $m = 20$ となり，全体に下側（表では上側）に偏ります。

x	n	e.p	Bin.p.	Lower p.	Upper p.	Central p.
16	100	0.2	0.0638	0.1923	0.8715	0.6824
17	100	0.2	0.0789	0.2712	0.8077	0.5466
18	100	0.2	0.0909	0.3621	0.7288	0.3828
19	100	0.2	0.0981	0.4602	0.6379	0.1974
20	100	0.2	0.0993	0.5595	0.5398	0.0993
21	100	0.2	0.0946	0.6540	0.4405	0.2919
22	100	0.2	0.0849	0.7389	0.3460	0.4677
23	100	0.2	0.0720	0.8109	0.2611	0.6186
24	100	0.2	0.0577	0.8686	0.1891	0.7401

上の表から，帰無仮説採択域を $x = 19, 20, 21$ ($b = 25\%$)，または $x = 18, 19, 20, 21, 22$ ($b = 50\%$)とします。期待確率(e.p.)が 0.5 ではないので，計算されたすべての確率(個別確率 Bin.p., 下側確率 Lower p. 上側確率 Upper p. 中央確率 Central p.)に对称性がありません。

以上で見たように，帰無仮説を肯定するときは，必然的にその危険率が高くなります(25, 50%)。しかし，危険率が 80%の帰無仮説と危険率が 20%の帰無仮説を区別なく棄却することは乱暴です。危険率が比較的小さな帰無仮説ならば，それを肯定することも可能だと思います。これは，ちょうど，危険率が 5%であっても帰無仮説を棄却して対立仮説を採択することが可能であることと似ています。違いは，帰無仮説を肯定するときの危険率がかかなり高い，ということです。帰無仮説を肯定するときも，それを棄却して対立仮説を採択するときも，常にその危険率を認識しておかなければなりません。有意水準だけを示すと，たとえば $p = 0.049$ も $p = 0.00001$ も同じ扱いをしていることになるからです。有意水準は便宜的に決めた目安にすぎません。

● プログラム

Function BinT(x, n, p, sel)'

‘二項分布確率, sel=0(現確率); sel=1(下側); 2(上側); 3(中心)

```

Dim i, Ar, s, e, m, t, c: ReDim Ar(n): Ar(0) = 0
For i = 1 To n
    Ar(i) = Ar(i - 1) + Log(i) '対数の階乗を配列に
Next i
If sel = 0 Then s = x: e = x 's=開始点 : e=終了点 sel=0(現確率)
If sel = 1 Then s = 0: e = x 's=開始点 : e=終了点 sel=1(下側累積確率)
If sel = 2 Then s = x: e = n 's=開始点 : e=終了点 sel=2(上側累積確率)
m = Application.RoundUp(n * p + p - 1, 0) '確率最大項
If sel = 3 And x < m Then s = x: e = n '中心確率
If sel = 3 And x >= m Then s = 0: e = x '中心確率
t = Int(Exp(Ar(n) - Ar(x) - Ar(n - x) + x * Log(p) + (n - x) * Log(1 - p)) * 10 ^ 7)
For i = s To e '累積確率
    c = Exp(Ar(n) - Ar(i) - Ar(n - i) + i * Log(p) + (n - i) * Log(1 - p))
    If sel < 3 Then BinT = BinT + c
    If sel = 3 And Int(c * 10 ^ 7) >= t Then BinT = BinT + c
Next
End Function

```

ここで用いた対数・指数変換(Log, Exp)については「確率」→「二項分布確率」を参照してください。

● Excel による実験

次のシートを用意します。

	A	B	C	D	E	F	X	Y
1	Binomial-test		Trial	x1	x2	x20	Sum	
2	Set		1	0	0	0	2	
3	p	0.2	2	0	0	0	1	
4	n	20	3	0	0	0	3	
5		x1-x20	4	1	0	0	4	
6			5	0	0	0	5	
7	Trial		6	0	0	0	2	
8	m	0.1945	7	0	0	0	6	
9	v	0.1567	8	1	0	0	5	
10			9	0	0	0	1	
11	Renew: [F9]		10	0	1	0	3	
12			11	0	0	1	3	

B 列に事前確率(p)とデータ数(n)を入れ、E2 に次の乱数を設定します。

$$E2: =IF(RAND()<B3,1,0)$$

上の式で[0, 1]の範囲の乱数が B3 で設定した事前確率(=0.2)よりも小さければ 1 を出力し、そうでなければ 0 を出力します。これを F:X までコピー

一し⁶、さらに E2:X2 の行を下に 2000 行コピーします。Y 列は E:X の和に
 します。

Y2:=SUM(E2:X2)

これを下に 2000 個コピーします。よって Y 列には[0, 20]の範囲で二項確
 率による頻度の分布ができます。[F9]キーを押すとこの実験用のデータが
 更新されます。

Z	AA	AB	AC	AD	AE
	Sum	P.trial	B.Dist.	B.cum	
	0	0.011	0.012	0.012	
	1	0.058	0.058	0.069	
	2	0.137	0.137	0.206	
	3	0.223	0.205	0.411	
	4	0.212	0.218	0.630	

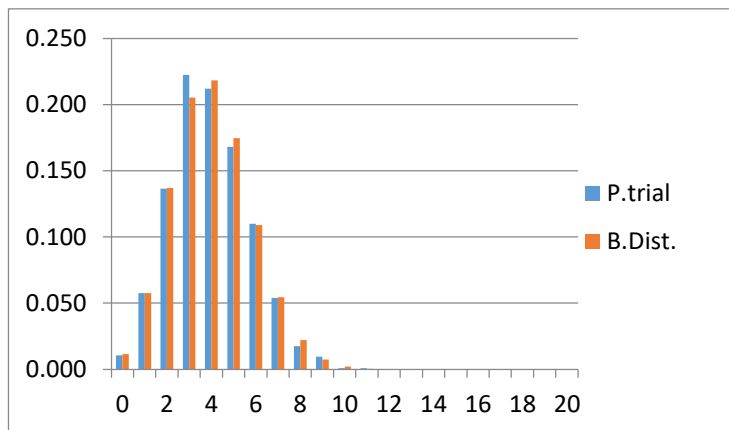
AA 列に[0,20]の連続数を入れ、AB2, AC2, AD2 に次を設定し、これらを
 下に 20 行コピーします。

AB2: =COUNTIF(Y:Y,"="&AA3)/2000

AC2: =BINOMDIST(AA2,\$B\$4,\$B\$3,0)

AD2: =BINOMDIST(AA2,\$B\$4,\$B\$3,1)

こうして AB 列に先に用意した乱数による実験で得られた数値を総数
 2000 で割った相対頻度が示され、AC 列に AA 列の頻度に対応する二項確
 率（個別確率）が示され、AD 列にその累積確率が示されます。次は AB
 列と AC 列の確率を比較するグラフです。一実験による確率の観測値と理
 論値が近似していることが確認できます。



⁶ 上の図では G:W(x3-x19)は非表示にしてあります。

9.1.2. 割合の有意性

事前確率(p)を 0.5, 個数 N=20, X=15 として, 次のシートを用意します。

	A	B	C	D	E
1	Data	x		Binomial test. Data	
2	1			p	0.5000
3	2	+		X	15
4	3	+		N	20
5	4	+		P	0.7500
6	5	+		Significance level	0.0500
7	6	+		Lower probability	0.9941
8	7	+		Lower critical value	6.0000
9	8	+		Lower significance	NS (H0)
10	9	+		Upper probability	0.0207
11	10	+		Upper critical value	14.0000
12	11	+		Upper significance	S. (H1)
13	12	+		Two-tailed probability	0.0414
14	13	+		Two-tailed critical value	14.0000
15	14	+		Two-tailed significance	S. (H1)
16	15	+			

Binomial test. Data

p 0.5

X =COUNTIF(B:B,"+")

N =COUNT(A:A)

P =E3/E4

Significance level 0.05

Lower probability =IF(E3=0,0,BINOMDIST(E3,E4,E2,1))

Lower critical value =CRITBINOM(E4,E2,E6)

Lower significance =IF(E7<E6,"S. (H1)", "NS (H0)")

Upper probability =IF(E3=0,1,1-BINOMDIST(E3-1,E4,E2,1))

Upper critical value =CRITBINOM(E4,E2,1-E6)

Upper significance =IF(E10<E6,"S. (H1)", "NS (H0)")

Two-tailed probability =2*MIN(E7,E10)

Two-tailed critical value =CRITBINOM(E4,E2,1-E6/2)

Two-tailed significance =IF(E13<E6,"S. (H1)", "NS (H0)")

上の上側確率(Upper probability)は, BINOMDIST が累積確率を返すので, 初めは, 出現回数(E4)から 1 を引いた数までの累積確率を計算し, その結果を 1 から引いた数値が, 当該の値以上のケースに対応する上側確率になります。出現回数(E4)が 0 の場合の上側確率は計算できませんが, これは必然的に 1 になります。

次のように設定すると頻度の二項検定を直接行うことができます。この

とき、事前確率(P)、データ数(N)、頻度(F)を直接書き込みます。たとえば、一般に左利きの人の割合が1割(0.100)として、20人の野球チームの中に4人の左利きの選手がいたとき(20%=4/20)、このチームの左利き選手が有意に多いかどうかを検定するには、次のように、直接セルに書き込んで、上側確率を求めます。

Frequency test. Result	
p	0.100
X	4
N	20
P	0.200
Significance level	0.050
Lower probability	0.957
Lower critical value	0.000
Lower significance	NS (H0)
Upper probability	0.133
Upper critical value	4.000
Upper significance	NS (H0)
Two-tailed probability	0.266
Two-tailed critical value	5.000
Two-tailed significance	NS (H0)

二項検定の結果、上側確率が.133 (> .05)なので、その数値に有意性がないことがわかります。

このように二項検定は、事前確率が分かっているとき、全数 N のうち生起頻度 X 回の割合 (%) の有意性を測ることができます。以下で扱うさまざまな検定法と違って、近似式を使わないため検定は正確です。そして、背景となる理論は二項確率だけなので理解が簡単です。そうしたことから言語研究に役立つ機会が多い方法だと言えるでしょう。

●ポアソン検定

大きな文書内(語数 N:きわめて大)で、小さな確率でランダムに生起する単語の生起数(X)が平均(M)を有意に下回るか、上回るかを検定します⁷。そして、その生起率(P)はコイン投げで「表」になる確率や(1/2)やサイコロを投げて「1」が出る確率(1/6)などと比べて一般にとっても小さい確率になります。このように多くの母数(N)のデータで小さな生起確率(P)で起こる現象の生起数(X)が有意に小さいか、または有意に大きいかを検定する方法と

⁷ たとえば定冠詞は名詞の前に現れるので、冠詞の出現が「ランダムである」とはいえませんが、ここではテキスト全体の定冠詞をマークしたときに、マークされた語の出現がランダムになっていることを見ます。

して「ポアソン検定」(Poisson test)が使われます。→「確率」「ポアソン分布」

このポアソン検定の帰無仮説と対立仮説は次のようになります。

帰無仮説：生起数(X)は平均(M)より有意に小さくはない。

対立仮説：生起数(X)は平均(M)より有意に小さい。

たとえば、生起率(P)が.000195 であるような現象が母数(N)が 20000 のとき、平均(M)は $P*N = 3.9$ になります。このとき、1度も生起しなかったとき(X:0), X=0 が平均と比べて統計的に有意に少ないといえるでしょうか？

そこで、はじめに有意率を.05 (5%)と設定し、「小さい」ほうのポアソン累積確率を計算すると.020 となります。この累積確率は有意率より小さいので、このようなことが起こるのは 5%以下、ということになり、5%の有意率（危険率）で帰無仮説が棄却され、対立仮説が採択されます。

PoissonLeft(0, 3.9) = .020 < a (.05), Significant

Poisson test: Left		
A	0.05	
P	0.000195	=195/10^6
N	20000	
M	3.900	=P*N
X	0	
Left P	0.020	=POISSON(X, M, TRUE)
Signif.	S	=IF(Left < A,"Yes","No")

次は、上と同じ現象が 8 回おきたときに、X=8 は M=3.9 より有意に大きいか否かを検定します。今回は「有意に大きいか否か」を検定するので、右側検定をします。そのときの帰無仮説と対立仮説は次のようになります。

帰無仮説：生起数(X)は平均(M)より有意に大きくはない。

対立仮説：生起数(X)は平均(M)より有意に大きい。

Poisson test: Right		
A	0.05	
P	0.000195	
N	20000	
M	3.900	
X	8	
Right P	0.0454015	=1-POISSON(X-1, M, TRUE)
Signif.	S	

上の結果から、帰無仮説が棄却され、対立仮説が採択されて、生起数(X)=8

は平均よりも有意に大きい，という結論を導きます。

$$\text{PoissonRight}(8, 3.9) = .045 < \alpha (.050), \text{Significant}$$

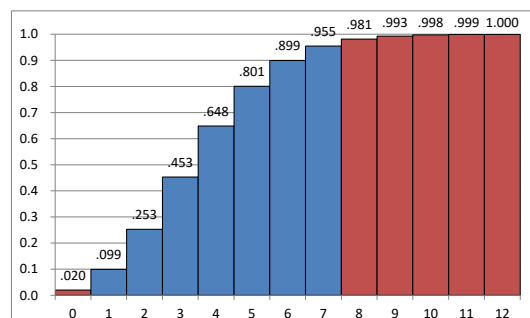
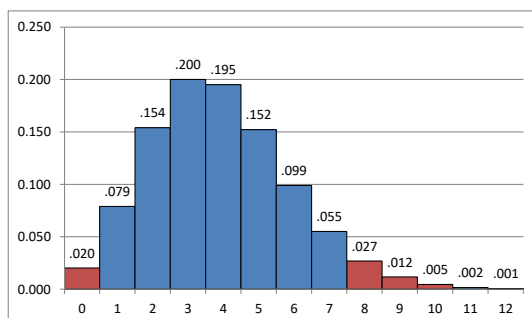
参考として $X=7$ の場合を検定すると右側の合計確率が.10 となるので有意性がなく，帰無仮説は棄却されません。

Poisson test: Right	
A	0.05
P	0.000195
N	20000
M	3.900
X	7
Right P	0.100517
Signif.	NS

$$\text{PoissonRight}(7, 3.9) = .100 > \alpha (.050), \text{No significant}$$

以上のことは次の表とグラフで確認できます。

X	Poisson.in.	Poisson.ac.
0	0.020	0.020
1	0.079	0.099
2	0.154	0.253
3	0.200	0.453
4	0.195	0.648
5	0.152	0.801
6	0.099	0.899
7	0.055	0.955
8	0.027	0.981
9	0.012	0.993
10	0.005	0.998
11	0.002	0.999
12	0.001	1.000



● 符号検定

次は二項確率を使う符号検定の場合です。AN:AO 列にデータを入力・コピーすると、次の設定によりすべての数値を自動的に計算します⁸。

J2: =IF(AN2>AO2, "+", "")

M4: =COUNTIF(AP:AP,"+")

G	H	I	J	K	L	M
Data	x	y	Sign		Sign test (+): x > y. Data	
1	18	11	+		p	0.500
2	22	12	+		X	10
3	25	35			N	15
4	12	22			P	0.667
5	22	18	+		Significance level	0.050
6	31	23	+		Lower probability	0.941
7	32	24	+		Lower critical value	4.000
8	22	22			Lower significance	NS (H0)
9	25	19	+		Upper probability	0.151
10	18	12	+		Upper critical value	11.000
11	20	33			Upper significance	NS (H0)
12	30	25	+		Two-tailed probability	0.302
13	29	22	+		Two-tailed critical value	11.000
14	27	21	+		Two-tailed significance	NS (H0)
15	25	40				

医薬品の効果，機械の性能，料理法の評価などに向上が見られたかどうかを統計的に検定する方法として「符号検定」(Sign test)が使われます。

たとえば外国語の新しい授業法 A (After)が，旧来の授業法 B (Before)と比較して，受講者のテストの成績に改善があったのかを検定することを考えてみましょう。受講者が 1 ~ 8 の計 8 名で，テストの成績が向上した人が 7 名であり，下降した人が 1 名であったとします。そのとき，向上した人をプラス(+)，下降した人をマイナス(-)とし，プラスの人数を Pos(itive)，マイナスの人数を Neg(ative)として，次のように数えます。プラスでもマイナスでもなかった人，つまり，A = B の人は以下の計算に含めません。そのとき全数(N)にも含めません。

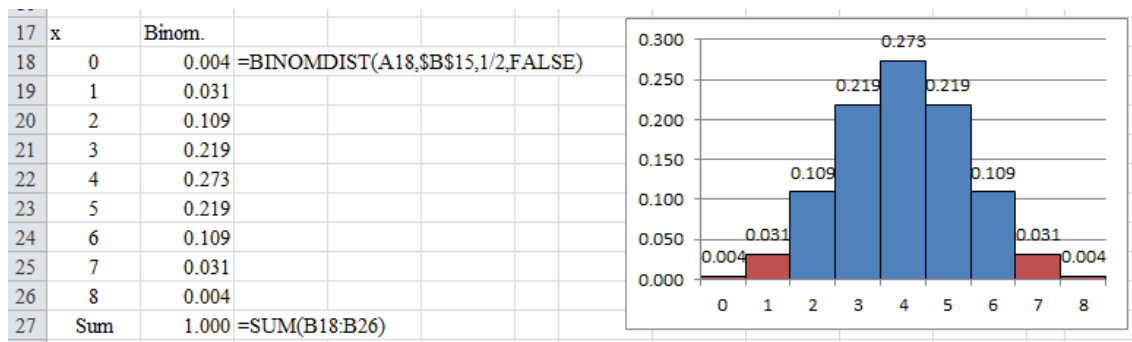
⁸ 直前の設定と異なる部分だけを示します。

	A	B	C	D	E	F	G	H	I	J	K	L
1	SIGN TEST		符号検定									
2												
3	(1) N <= 25		(small sample)									
4												
5	N	A	B	A - B	Sign		Pos.	7	=COUNTIF(E6:E13,"+")			
6	1	3	1	2	+		Neg.	1	=COUNTIF(E6:E13,"-")			
7	2	4	3	1	+		Side	Left				
8	3	6	1	5	+		Alpha	0.05				
9	4	2	4	-2	-		N	8	=SUM(H5:H6)			
10	5	9	7	2	+		X	1	=MIN(H5:H6)			
11	6	5	3	2	+		Bin.ac.	0.035	=BINOMDIST(H10,H9,1/2,TRUE)			
12	7	6	2	4	+		Signific.	Yes	=IF(H11<H8,"Yes","No")			
13	8	10	8	2	+							

このときの帰無仮説は「プラス数はマイナス数と比べて有意な差がない」とし、対立仮説は「プラス数はマイナス数と比べて有意な差がある」とします。はじめに有意確率(Alpha)を 5% (または 1%) として定め、次に、プラスの人数(Pos)とマイナスの人数(Neg)を比べて、少ない方を X として、1/2 の確率で起こることが、N 回の試行のうち、成功数が X であったときの二項分布累積確率(Bin.ac.)を求め、これがはじめに決めておいた有意確率以下であれば、帰無仮説を棄却し (プラス数はマイナス数と比べて有意な差がない、とは言えない)、対立仮説を採択します (Bin.ac.(1, 7, 1/2) = 0.035 < 0.05)。

実際に上図のように、Excel 関数を使って計算してみると二項分布累積確率は 3.5%になるので、対立仮説「プラス数はマイナス数と比べて有意な差がある」が採択されます。

さて、この検定ではプラスの人数(=7)を取り上げているのに、マイナスの人数(=1)で、左側の棄却域で検定しているのはなぜでしょうか？実は、プラスの人数(7)で右側の棄却域で検定しても、次の図が示すように、確率 1/2 の二項分布は左右対称になるので、同じ結果になるからです。



ここで左側の棄却域の確率の和は $0.004 + 0.031 = 0.035$ (3.5%)であり、右側の確率の和も $0.031 + 0.004 = 0.035$ (3.5%)になっています。しかし、左側の合計確率は $X = 1, 2$ の累積確率をそのまま使えばよいのですが、右側ではプラスの人数(=7)までの累積確率($X = 1, 2, \dots, 7$)を使うわけにはいきま

せん。棄却域は $X = 7, 8$ の領域ですから（上図），次のように全体の確率 (=1.000)から $7 - 1 = 6$ までの累積確率を引かなければならないのです。

Pos.	7
Neg.	1
Side	Right
Alpha	0.05
N	8
X	7
Bin.ac.	$N11 = 0.035$
Signific.	Yes

$$N11=1-BINOMDIST(N10-1,N9,1/2,TRUE)$$

そこで，符号検定では左側棄却域の確率を直接求めるために，プラスの数とマイナスの数の少ない方(=MIN)を選択して，これを X 値とします。

ここで，いくつかのデータ例の符号検査を比較してみましょう。

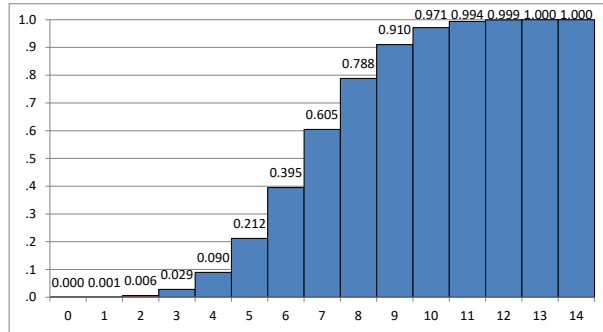
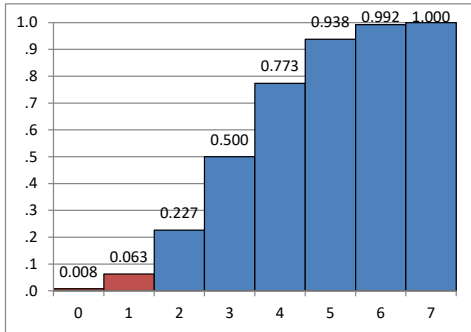
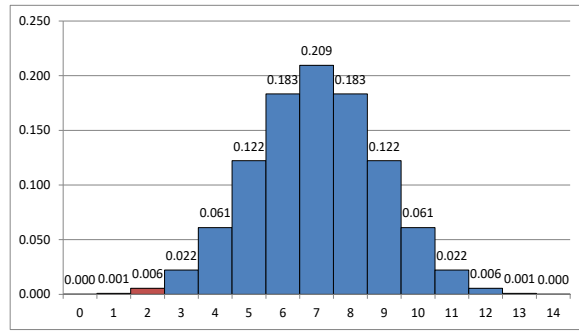
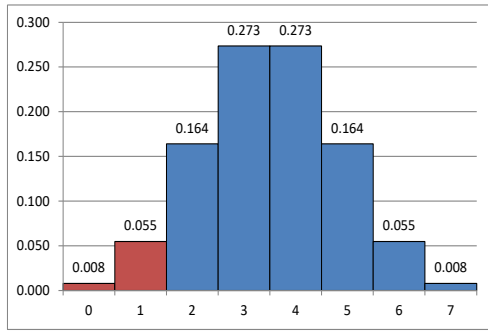
Pos.	6	Pos.	6	Pos.	12
Neg.	2	Neg.	1	Neg.	2
Side	Left	Side	Left	Side	Left
Alpha	0.05	Alpha	0.05	Alpha	0.05
N	8	N	7	N	14
X	2	X	1	X	2
Bin.ac.	0.145	Bin.ac.	0.063	Bin.ac.	0.006
Signific.	No	Signific.	No	Signific.	Yes

左の表はプラスとマイナスの値が 6 と 2 のときの符号検査の結果を示します。このとき，累積確率が 0.145 (14.5%)になるので，帰無仮説は棄却できません。つまり，プラス数はマイナス数と比べて有意な差がある，とは言えません ($\text{Bin.ac.}(2, 8, 1/2) = 0.147 > 0.05$)。

中の表ではデータ数(N)が 7 の場合を示しますが，累積確率は 0.063 (6.3%)ですから，有意確率 0.05 以下にはならないので，やはり帰無仮説を棄却できません($\text{Bin.ac.}(1, 7, 1/2) = 0.063 > 0.05$)。

右の表は $N=14, X=2$ の場合です。これは表中で見た $N=7, X=1$ の 2 倍の規模になっています。このときの累積確率は 0.006 (0.6%)なので，有意確率以下になり，帰無仮説が棄却され，対立仮説が採用されます($\text{Bin.ac.}(2, 14, 1/2) = 0.006 < 0.05$)。

このことは一見不思議に思えるかもしれませんが。X/N の比率はどちらも同じなのに，サンプル数(N)が大きくなると有意な差が生まれることが多いのです。この理由は次の 4 つの図（上が個別の確率，下が累積確率）を比較するとわかります。



左図を見ると、8 つだけの確率の分布を示しているので分類が粗く、最初のいくつかの確率を足し上げると、すぐに大きな数値になることがわかります。一方、右図では 15 の確率があり、細かな間隔で数値が上昇していくので、すぐに大きな値になることはありません。サンプル数が小さいとき、検定をしないで、安易な判断をすることはできません。

■ スペイン語の文字列 *cua*

現代スペイン語の単語ではヨーロッパの他の言語にある文字列 *qua* を使わず、代わりに *cua* が使われます(*cuarto, cuantitativo, ...*)。しかし歴史的には *qua* と *cua* が競合していました。次の表は 1200 年代から 100 年ごとに収集した公証文書に見られる *qua* を含む語(=*qua*=)と *cua* を含む語(=*cua*=)の頻度、総語数(N)、全体の確率($P = \text{Sum}/N$)を示します。

FA	1200	1300	1400	1500	1600	1700	Sum	P
= <i>qua</i> =	461	421	1285	2432	1604	688	6891	0.005108
= <i>cua</i> =	64	109	41	33	43	53	343	0.000254
N	234003	240786	275208	291731	169981	137282	1348991	

そこで、たとえば 1600 年代と 1700 年代の=*cua*=の生起数(43, 53)について、それらが平均より有意に大きいか否かを検定します。

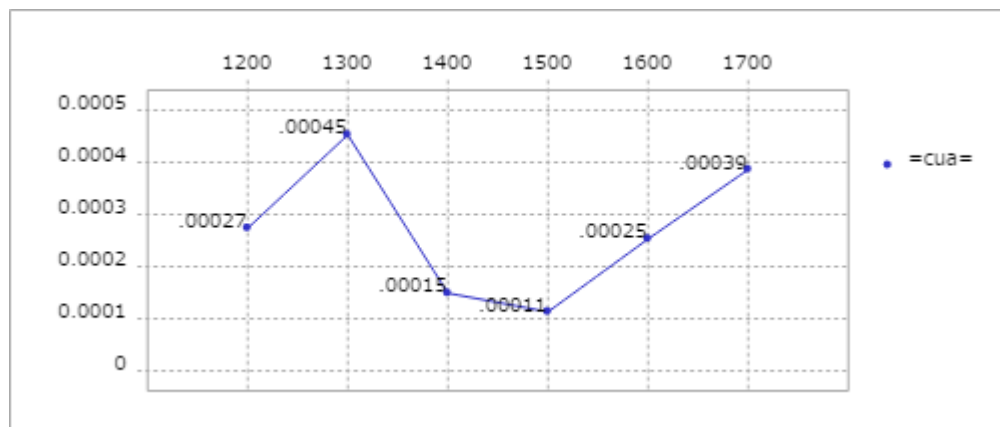
ここで、たとえば 1600 年代と 1700 年代の=*cua*=の生起数(43, 53)について、それらが平均より有意に大きいか否かを検定します⁹。

⁹文書内でランダムに表れる語のように確率(P)が小さく母数(N)が大きな頻度データはポアソン分布に従っていると考えられるので、一般にポアソン検定が行われます。しかし、「確率」のセクションで見たように、ポアソ

Test	1200	1300	1400	1500	1600	1700
A	0.05	0.05	0.05	0.05	0.05	0.05
P	0.000254	0.000254	0.000254	0.000254	0.000254	0.000254
N	234003	240786	275208	291731	169981	137282
M	59.499	61.223	69.976	74.177	43.220	34.906
X	64	109	41	33	43	53
Upper p.	0.2937	0.0000	0.9999	1.0000	0.5309	0.0025
Signif.	NS	S	NS	NS	NS	S

上の表に見られるように、1600年代では有意差はありませんが(NS: No significant), 1700年代になって有意に cua が平均を上回っていることがわかります(S: Significant)。=cua=は1300年代にも有意な生起率を記録しています。

次は=cua=の各年代における生起率(生起数/総語数)の変化を示します。



以上の結果から=cua=の生起率が1600年代から1700年代にかけて上昇していますが、それが有意に平均値を超えたのは1700年代であったことがわかります。その上昇は突然生じたのではなく1300年代に先行する高頻度の使用が潜在的に影響した、と考えます。

一般に検定はたとえば有意率(危険率)が5%であれば、20回に1回は判定を間違える可能性があるので複数回行ってはいけない、といわれます(検定の多重性)。しかし、上の分析では6回の検定を行っていますが、それぞれが独立した検定であると考えます。検定の多重性は、たとえば1700年代の文書をさまざまな資料から複数回収集して同じ検定を繰り返す、という作業で生じる問題です。そのとき、有意差が現れるまで検定を繰り返す、というのでは検定の意味がありません。

ン確率分布は二項確率分布の近似値なので、やや誤差が見られます。それでもポアソン確率分布が使われるのは、大きな階乗や指数の計算が困難なためですが、これは対数・指数変換をすれば解決できるので(→「確率」), それを使った二項分布確率で計算します。

9.2. 平均値の Z 検定 : 1 標本

標本の平均値の有意性を検定するためには、Z 検定 (Z-test: 標本数が 30 以上) と t 検定 (t-test: 標本数が 30 未満) が使われます。はじめに Z 検定を扱います。

9.2.1. 標本 Z 値の確率分布

次の母集団の平均(m)と標本平均(M)の差について次の定理があります。

平均 m 、分散 v の任意の母集団からサイズ(n)を多くしていくと、抽出した標本の平均(M)を含む次の式 Z の分布は次第に標準正規分布 $N(0, 1)$ に近づく。

$$Z = (M - m) / (v / n)^{1/2} \sim N(0, 1)$$

この定理が実現される様子を次の Excel シートで実験して確かめます。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	Z-dist		Trial	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	x17	x18	x19	x20	Mean	Z	
2	Set		1	8.2	5.9	8.2	12.9	9.8	15.2	10.6	9.2	9.3	4.5	11.5	9.8	9.6	5.6	10.6	7.1	11.5	10.9	10.1	3.8	9.218	-1.090	
3	m	10	2	11.5	12.8	8.6	8.2	9.5	12.2	15.2	12.4	11.9	7.2	15.8	9.4	11.6	7.8	10.4	11.4	14.3	13.7	2.5	7.1	10.678	0.972	
4	v	10	3	17.0	4.6	6.7	12.1	16.2	13.1	6.6	10.4	7.1	6.3	7.3	9.3	9.0	11.8	13.2	7.0	10.7	6.7	11.1	11.4	9.888	-0.144	
5	sd	3.162	4	6.5	9.9	13.9	14.1	7.4	15.2	6.6	11.5	9.3	13.0	13.7	8.7	9.6	13.3	12.4	5.3	14.8	10.9	12.0	7.8	10.792	1.134	
6	n	20	5	10.4	9.6	11.3	13.3	13.5	9.2	16.2	10.8	8.2	12.3	9.6	9.4	7.6	7.9	14.4	9.0	6.9	8.0	9.9	12.4	10.503	0.725	
7		x1-x20	6	9.2	12.0	11.7	9.0	10.1	7.5	6.5	8.4	8.2	9.1	9.9	14.3	10.6	11.2	11.3	6.9	12.8	6.8	10.0	6.4	9.598	-0.554	
8			7	10.7	10.9	13.3	10.1	6.1	15.6	7.3	11.6	12.4	12.1	11.8	11.4	10.6	5.7	8.4	8.6	9.4	17.3	11.0	5.8	10.509	0.733	
9	Trial		8	14.1	12.2	6.0	9.4	11.8	6.0	10.8	9.3	6.9	11.5	8.7	13.4	9.0	9.9	8.2	7.6	7.9	9.3	7.8	14.6	9.715	-0.387	
10	M	9.990	9	11.6	9.7	8.2	9.1	6.2	11.3	13.5	11.1	11.3	12.2	15.6	6.2	10.1	12.4	9.7	11.3	8.8	11.5	12.4	8.3	10.530	0.764	
11	V	10.025	10	9.3	6.9	10.5	14.5	12.5	9.6	7.8	10.0	9.6	9.8	10.9	7.4	6.0	9.3	9.1	9.4	1.8	4.5	7.2	15.4	9.077	-1.288	
12	SD	3.166	11	9.7	8.5	12.0	9.5	11.8	13.2	5.7	13.4	8.8	11.1	10.3	7.8	6.7	8.0	10.3	9.7	9.0	9.9	13.8	6.1	9.756	-0.330	
13			12	9.5	7.1	11.8	7.1	7.2	12.7	13.1	10.8	8.3	6.6	13.0	8.4	7.3	7.9	12.2	8.7	11.8	15.9	11.4	7.8	9.924	-0.092	
14	Renew: [F9]		13	7.3	13.4	8.5	12.4	14.3	11.8	11.5	9.2	9.3	6.8	10.4	12.8	10.4	10.2	8.8	16.8	6.8	13.0	12.1	10.8	10.833	1.192	

上の B3 と B4 に正規分布乱数の平均(m)と分散(v)を入れます。E2 に次の関数を設定します¹⁰。

$$E2: =\text{NORMINV}(\text{RAND}(), \$B\$3, \text{SQRT}(\$B\$4))$$

これを F:X 列にコピーして、その後 E2:X2 までの行を下に 2000 行コピーします。B9 と B10 に上の式で生成された 20*2000 = 40000 個の乱数の平均(m)と分散(v)が表示されます。

$$B9: =\text{AVERAGE}(E:X) \leftarrow m$$

$$B10: =\text{VARP}(E:X) \leftarrow v$$

$$B11: =\text{SQRT}(B10) \leftarrow sd$$

¹⁰ 先の定理は「任意の母集団」について成立するので、正規分布乱数でなくても、乱数そのもの、たとえば $\text{RAND}()$ でもかまいません。しかし、そうすると母集団に一様分布を想定することになり不自然です。

この m と v を母集団の平均(m)と分散(v)とします。

Y2 と Z2 にサイズ(n)=20 の標本の平均と, 上の定理の Z 値を設定します。

Y2: =SUM(E2:X2)/20 ←M

Z2: =(Y2-\$B\$9)/SQRT(\$B\$10/20) ←Z

この Y2:Z2 を下に 2000 個コピーします。これらをそれぞれの標本の平均(M)と Z 値(Z)とします。

次に AB:AG に 0,5 幅の Z 値にそれぞれ対応する実験で得られた頻度による確率(P.trial), Excel 関数による個別確率(N.dist), その累積確率(N.cum) をセットします。

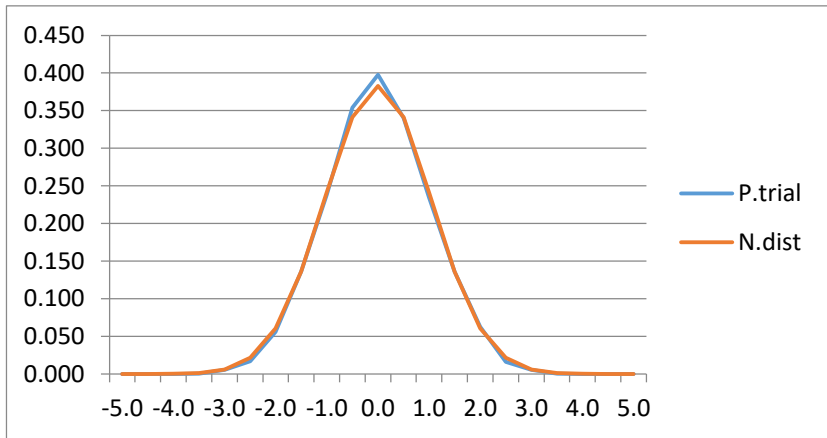
Z	from	to	P.trial	N.dist	N.cum
-5.0	-5.5	-4.5	0.000	0.000	0.000
-4.5	-5.0	-4.0	0.000	0.000	0.000
-4.0	-4.5	-3.5	0.000	0.000	0.000
-3.5	-4.0	-3.0	0.001	0.001	0.000
-3.0	-3.5	-2.5	0.005	0.006	0.001
-2.5	-3.0	-2.0	0.017	0.021	0.006
-2.0	-2.5	-1.5	0.061	0.061	0.023
-1.5	-2.0	-1.0	0.145	0.136	0.067
-1.0	-1.5	-0.5	0.236	0.242	0.159
-0.5	-1.0	0.0	0.343	0.341	0.309
0.0	-0.5	0.5	0.389	0.383	0.500
0.5	0.0	1.0	0.337	0.341	0.691
1.0	0.5	1.5	0.246	0.242	0.841
1.5	1.0	2.0	0.136	0.136	0.933
2.0	1.5	2.5	0.057	0.061	0.977
2.5	2.0	3.0	0.023	0.021	0.994
3.0	2.5	3.5	0.007	0.006	0.999
3.5	3.0	4.0	0.001	0.001	1.000
4.0	3.5	4.5	0.001	0.000	1.000
4.5	4.0	5.0	0.000	0.000	1.000
5.0	4.5	5.5	0.000	0.000	1.000

AE2: =COUNTIFS(\$Z2:\$Z\$2001,">"&AC2,\$Z2:\$Z\$2001,"<="&AD2)/2000

AF2: =NORMSDIST(AD2)-NORMSDIST(AC2)

AG2: =NORMSDIST(AB2)

次が実験で得られた標本で計算された確率(P.trial)と, Excel 関数による標準正規分布 N(0, 1)の個別の確率(N.dist)を示すグラフです。



上の表とグラフを見ると実験で得られた頻度による確率(P.trial)と Excel 関数による個別確率がかなり近似していることがわかります。

9.2.2. Z 値の有意性

Z 値が標準正規分布 $N(0, 1)$ に従うことを使って、母平均の区間推定と母平均と標本平均の有意性の検定(Z 検定)をします。

	A	B	C	D	E
1	Data	x		Z-test (m:M). Data	
2	1	18		m	21.000
3	2	22		v	45.000
4	3	25		sd	6.708
5	4	12		N	30
6	5	22		M	23.233
7	6	31		Significance level	0.050
8	7	32		Z-value	1.824
9	8	22		Lower probability	0.966
10	9	25		Lower critical value	-1.645
11	10	18		Lower significance	NS (H0)
12	11	20		Upper probability	0.034
13	12	30		Upper critical value	1.645
14	13	29		Upper significance	S. (H1)
15	14	27		Two-tailed probability	0.068
16	15	25		Two-tailed critical value	1.960
17	16	11		Two-tailed significance	NS (H0)

AM 列の下の項目にそれぞれ対応する関数をセットします。

Z-test (m:M). Data
 m 21
 v 45
 sd =SQRT(E3)

N =COUNT(B:B)
 M =AVERAGE(B:B)
 Significance level 0.05
 Z-value =(E6-E2)/SQRT(E3/E5)
 Lower probability =NORMSDIST(E8)
 Lower critical value =NORMSINV(E7)
 Lower significance =IF(E9<E7,"S. (H1)", "NS (H0)")
 Upper probability =1-E9
 Upper critical value =NORMSINV(1-E7)
 Upper significance =IF(E12<E7,"S. (H1)", "NS (H0)")
 Two-tailed probability =2*MIN(E9,E12)
 Two-tailed critical value =NORMSINV(1-E7/2)
 Two-tailed significance =IF(E15<E7,"S. (H1)", "NS (H0)")

青色のセルに母集団の平均(m), 分散(v), そして検定の有意水準を入れます。そして AJ 列に頻度データをセットすると, AM の列に Z 検定の結果が表示されます。

有意水準=危険率を 1%にするときには係数 0.5%点と 99.5%点(-2.58, 2.58)を使います。

Z-value は標本平均(M), 母分散(v), サンプルサイズ(n)によって求めます。

$$Z = (M-m) / (v/n)^{1/2}$$

「下側確率」(Lower probability)は Z 値に対応する正規分布確率です。「下側閾値」(Lower critical value)は有意水準に対応する Z 値です。「下側有意性」(Lower significance)では下側確率が有意水準以下であったときに「有意」(S. H1)として対立仮説(H1)を採用します。そうでなければ「有意でない」(NS. H0)として帰無仮説 H0 を棄却しません。

同じことを上側でも行います。

以上は「片側検定」(one-sided test)であり, 標本平均(M)が母平均(m)より有意に少ないか(下側検定), または有意に多いか(上側検定)を検定する場合です。

標本平均(M)が母平均(m)と有意に違うか否かを検定するときは「両側検定」(Two-tailed test)を行います。そのときは, 下側確率と上側確率のうち小さな方が, 有意水準の半分よりも小さなきに有意と判定されます。両側を合わせて有意水準と比較するので, Z 値の確率は有意水準の半分以下でなければ有意にならない, と判断されます。つまり, 基準が厳しくなるのです。

9.2.3. Z 検定の例

たとえば、旧来の外国語教授法で行った授業の受講者の期末テストの平均点が 82 点、その分散が 92 であったとき、今回新しい授業方法に変えて、平均点が 85 点に上昇したケースについて、この成績アップが統計的に有意であったかどうかを検定します。履修者数は 32 名でした¹¹。

この検定での帰無仮説は「成績は向上していない」ということになるでしょう。ここでは「成績が向上したこと」（対立仮説）を検定するので、片側検定で上側確率(upper probability)を見ます。

Z-test (m:M)	
m	82.000
v	92.000
N	32
M	85.000
Significance level	0.050
Z-value	1.769
Lower probability	0.962
Lower critical value	-1.645
Lower significance	NS (H0)
Upper probability	0.038
Upper critical value	1.645
Upper significance	S. (H1)
Two-tailed probability	0.077
Two-tailed critical value	1.960
Two-tailed significance	NS (H0)

その結果、Z-value=1.769 > Z.crit.val.(5%):1.645 で帰無仮説が棄却され(p.=3.8%)、「成績は向上した」という対立仮説が採用されます。

9.3. 平均値の Z 検定 : 2 標本

9.3.1. 標本 Z 値の確率分布

標本数が大きな 2 つのデータ(D1, D2)の平均値の差について次の式で示される Z 値を求めると、その Z 値が標準正規分布にしたがうことが知られています。

$$Z2 = (M1 - M2) / (V1/N1 + V2/N2)^{1/2}$$

¹¹ Z 検定では母集団は正規分布に従うことがわかっていることが理想的ですが、母集団が正規分布に従うかどうかわからなくても、用いる標本のサイズが十分大きければ（一般に 30 以上ならば）よい、と考えられます。

ここで、M1, M2 は D1, D2 の平均値、V1, V2 は D1, D2 の不偏分散、N1, N2 は D1, D2 の標本数を示します。

このことを Excel を使って実験して確かめます。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	Z-dist			Trial	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	x17	x18	x19	x20	x.Mean	x.u.Var
2	Set			1	7.5	9.8	13.0	13.6	6.5	11.5	7.2	5.6	11.1	10.6	7.8	9.4	13.4	13.5	8.2	8.5	8.6	5.9	7.6	9.2	9.435	6.549
3	m	10		2	11.5	17.7	9.8	9.3	7.9	9.2	7.4	8.7	5.4	11.9	5.7	10.5	12.2	10.3	11.4	7.0	18.5	11.6	10.4	12.8	10.466	11.228
4	v	10		3	9.6	13.1	10.2	9.0	9.4	10.1	11.5	7.5	9.2	10.4	14.2	13.7	3.3	10.9	11.8	3.5	10.5	10.6	12.9	14.8	10.311	9.166
5	sd	3.162		4	9.8	13.0	9.9	6.0	14.5	12.5	9.9	7.9	10.1	15.4	9.7	11.1	8.6	6.1	9.1	5.3	6.6	12.8	4.4	8.1	9.521	9.345
6	n	20		5	11.2	11.5	10.9	9.5	6.2	6.8	4.8	10.6	12.4	15.4	8.8	7.9	4.2	4.0	11.9	12.4	8.2	9.1	5.5	7.5	8.931	9.678

AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW
y1	y2	y3	y4	y5	y6	y7	y8	y9	y10	y11	y12	y13	y14	y15	y16	y17	y18	y19	y20	y.Mean	y.u.Var	Z
10.9	7.8	13.6	4.5	9.7	11.3	13.5	6.4	11.3	12.9	6.5	8.7	7.0	16.7	11.1	9.1	6.5	13.4	7.9	21.3	10.512	16.210	-1.009
10.5	10.4	8.3	7.3	10.5	9.9	8.3	11.5	4.9	7.6	5.1	11.6	7.5	12.0	12.9	14.2	3.2	11.2	7.2	11.1	9.261	8.255	1.221
10.0	7.9	11.5	5.8	13.9	9.6	9.0	14.1	15.8	15.1	4.3	12.0	4.0	7.3	14.0	11.8	8.2	10.7	1.7	10.2	9.857	15.167	0.412
11.3	10.0	10.8	9.4	7.2	7.2	3.8	8.3	7.7	11.2	9.3	9.8	11.4	16.0	9.7	15.6	12.2	11.5	8.6	7.8	9.945	7.930	-0.456
7.7	10.0	13.1	4.1	9.9	10.3	14.9	10.5	14.5	9.7	4.8	3.9	14.8	10.7	8.5	10.2	11.9	8.1	7.7	10.4	9.772	10.337	-0.841
8.3	13.1	12.8	7.9	10.9	11.3	9.6	10.3	9.2	7.5	9.8	9.7	11.0	10.4	7.0	4.7	6.9	8.1	6.6	13.4	9.423	5.456	-0.811

E2: =NORMINV(RAND(), \$B\$3, SQRT(\$B\$4))

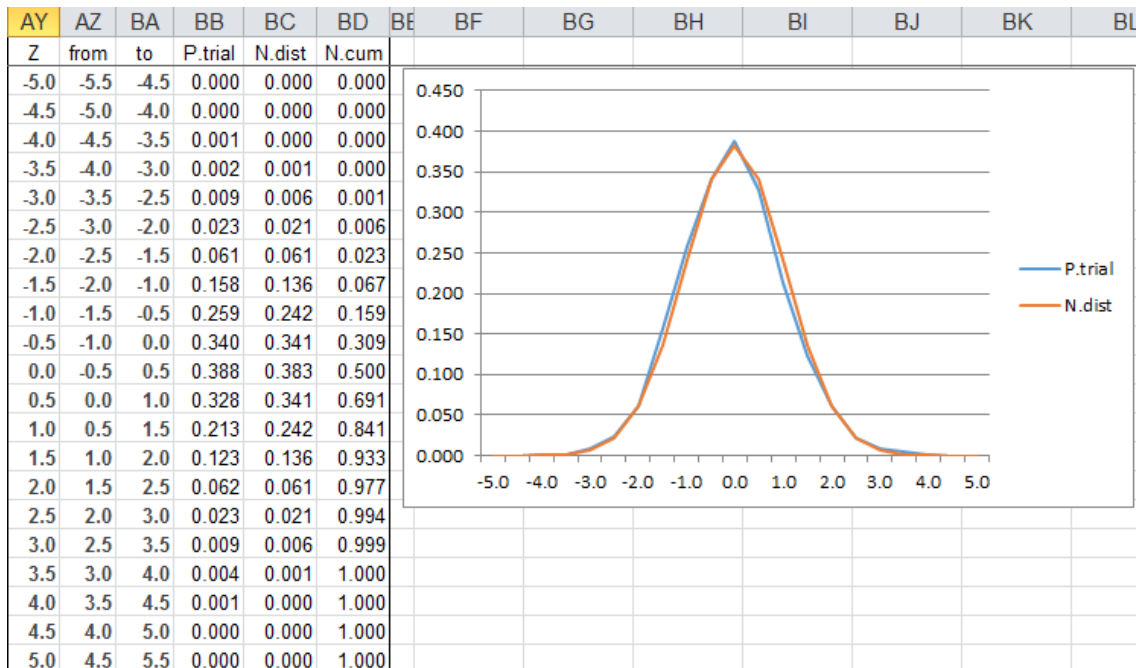
Y2: =AVERAGE(E2:X2)

Z2: =VAR(E2:X2)

AU2: =AVERAGE(AA2:AT2)

AV2: =VAR(AA2:AT2)

AW2: =(Y2-AU2)/SQRT(Z2/20+AV2/20)



BB2:

=COUNTIFS(\$AW\$2:\$AW\$2001,">"&AZ2,\$AW\$2:\$AW\$2001,"<="&BA2)/2000

BC2: =NORMSDIST(BA2)-NORMSDIST(AZ2)

BD2: =NORMSDIST(AY2)

たしかに上の表とグラフを見ると，Z 値が標準正規分布にほぼ一致していることがわかります。よって，2 つの標本の平均値の差の有意性を，Z 値を介して，Z 値に対応する確率から求めることができます。

9.3.2. Z 値の有意性

次は，2 標本の平均差の有意性を求める Z 検定の結果です。ABC 列はデータ例です。ここでは外国語テストの点数で，練習前(Before)と練習後(After)の結果を比較します。帰無仮説(H0)は「練習の前と後を比較して，得点が上昇しない」，対立仮説(H1)は「得点が上昇する」となります。よって，片側検定になります。有意水準として 0.05 です。中図では Excel 関数を使っています。右図は Excel の「データ分析」の出力です。

	A	B	C	D	E	F	G	H	I	J	K
1	標本数	After	Before		Z 2-test (m:M). Data	After	Before		z-検定: 2 標本による平均の検定		
2	1	19	19		M	22.4000	19.0333				
3	2	30	20		u.V	25.3517	20.1713			After	Before
4	3	29	15		u.SD	5.0350	4.4912		平均	22.4000	19.0333
5	4	27	23		N	30.0000	30.0000		既知の分散	25.3517	20.1713
6	5	18	16		Significance level	0.0500			観測数	30.0000	30.0000
7	6	20	13		Z-value	2.7330			仮説平均との差異	0.0000	
8	7	35	14		Lower probability	0.9969			z	2.7330	
9	8	17	27		Lower critical value	-1.6449			P(Z<=z) 片側	0.0031	
10	9	14	24		Lower significance	NS (H0)			z 境界値 片側	1.6449	
11	10	24	30		Upper probability	0.0031			P(Z<=z) 両側	0.0063	
12	11	26	12		Upper critical value	1.6449			z 境界値 両側	1.9600	
13	12	21	16		Upper significance	S. (H1)					
14	13	24	20		Two-tailed probability	0.0063					
15	14	24	16		Two-tailed critical value	1.9600					
16	15	15	22		Two-tailed significance	S. (H1)					

F2: =AVERAGE(B:B), G2: =AVERAGE(C:C)

F3: =VAR(B:B), G3: =VAR(C:C)

F4: =STDEV(B:B), G4: =STDEV(C:C)

F5: =COUNT(B:B), G5: =COUNT(C:C)

F7: =(F2-G2)/SQRT(F3/F5+G3/G5)

F8: =NORMSDIST(F7)

F9: =NORMSINV(F6)

F10: =IF(F8<F6,"S. (H1)", "NS (H0)")

F11: =1-F8

F12: =NORMSINV(1-F6)

F13: =IF(F11<F6,"S. (H1)", "NS (H0)")

F14: =2*MIN(F8,F11)

F15: =NORMSINV(1-F6/2)

F16: =IF(F14<F6,"S. (H1)", "NS (H0)")

この結果，上側確率が.003 になるので，得点の上昇(19.0 → 22.4)が有意であったことがわかります。

9.4. 平均値の t 検定

9.4.1. t 値の確率分布

標本のサイズが小さい(n=30 以下) 標本の平均の有意性を検定するには、標準正規分布を利用した Z 検定は使えません。その代わりに t 分布を使う t 検定を行います。はじめに、t 分布の性質を次の Excel シートで実験して確かめます。

	A	B	C	D	E	F	G	H	I	J	K
1	t.0.dist			Trial	x1	x2	x3	x4	Mean	u.Var.	t
2	Set			1	11.7	11.3	10.6	8.9	10.6	1.6	0.978
3	m	10.000		2	10.5	10.8	8.6	11.9	10.4	2.0	0.612
4	v	3.000		3	9.8	9.5	11.1	7.8	9.6	1.8	-0.660
5	sd	1.732		4	9.6	9.3	12.2	7.3	9.6	4.0	-0.432
6	n	4		5	9.2	9.4	10.8	8.2	9.4	1.1	-1.185
7				6	10.9	10.6	10.5	6.9	9.7	3.5	-0.333
8	Trial			7	12.8	14.5	8.1	10.8	11.5	7.6	1.110
9	M	10.016		8	8.1	10.8	10.8	11.6	10.3	2.3	0.422
10	V	2.991		9	7.8	9.9	8.1	11.2	9.2	2.7	-0.946
11	SD	1.730		10	7.7	11.4	11.1	12.8	10.7	4.6	0.678
12				11	13.3	6.2	10.1	10.6	10.1	8.7	0.038
13	Renew: [F9]			12	12.6	12.5	10.8	8.2	11.0	4.2	0.983
14				13	8.2	8.6	8.2	10.4	8.4	0.0	1.044

先の Z 値の確率分布と同様にして、上の B3 と B4 に正規分布乱数の平均(m)と分散(v)を入れます。E2 に次の関数を設定します。

E2: =NORMINV(RAND(), \$B\$3, SQRT(\$B\$4))

これを F:H 列にコピーして、その後 E2:H2 までの行を下に 2000 行コピーします。B9 と B10 に上の式で生成された 4*2000 = 8000 個の乱数の平均(m)と分散(v)が表示されます。

B9: =AVERAGE(E:H) ← m

B10: =VARP(E:H) ← uv

B11: =SQRT(B10) ← usd

この m と v を母集団の平均(m)と分散(v)とします。

I2, J2, K2 に、標本の平均、分散、t 値をそれぞれ設定します。

I2: =AVERAGE(E2:H2) ← Mean

J2: =VAR(E2:H2) ← u.Var.

K2: =(I2-\$B\$9)/SQRT(J2/4) ← t 値

この E2:K22 を下に 2000 個コピーします。これらをそれぞれの標本の平均(Mean), 不偏分散(u.Var), t 値(t)とします。

M	N	O	P	Q	R
t	from	to	P.trial	P.dist.	P.cum.
-5.0	-5.5	-4.5	0.005	0.004	0.008
-4.5	-5.0	-4.0	0.009	0.006	0.010
-4.0	-4.5	-3.5	0.011	0.009	0.014
-3.5	-4.0	-3.0	0.012	0.015	0.020
-3.0	-3.5	-2.5	0.022	0.024	0.029
-2.5	-3.0	-2.0	0.044	0.041	0.044
-2.0	-2.5	-1.5	0.080	0.071	0.070
-1.5	-2.0	-1.0	0.128	0.126	0.115
-1.0	-1.5	-0.5	0.194	0.210	0.196
-0.5	-1.0	0.0	0.291	0.304	0.326
0.0	-0.5	0.5	0.359	0.349	0.500
0.5	0.0	1.0	0.321	0.304	0.674
1.0	0.5	1.5	0.211	0.210	0.804
1.5	1.0	2.0	0.119	0.126	0.885
2.0	1.5	2.5	0.073	0.071	0.930
2.5	2.0	3.0	0.043	0.041	0.956
3.0	2.5	3.5	0.027	0.024	0.971
3.5	3.0	4.0	0.017	0.015	0.980
4.0	3.5	4.5	0.007	0.009	0.986
4.5	4.0	5.0	0.004	0.006	0.990
5.0	4.5	5.5	0.003	0.004	0.992

次に、M, N, O, P, Q, R 列をそれぞれ設定します。

N2: =M2-0.5

O2: =M2+0.5

P2: =COUNTIFS(\$K\$2:\$K\$2001,">"&N2,\$K\$2:\$K\$2001,"<="&O2)/2000

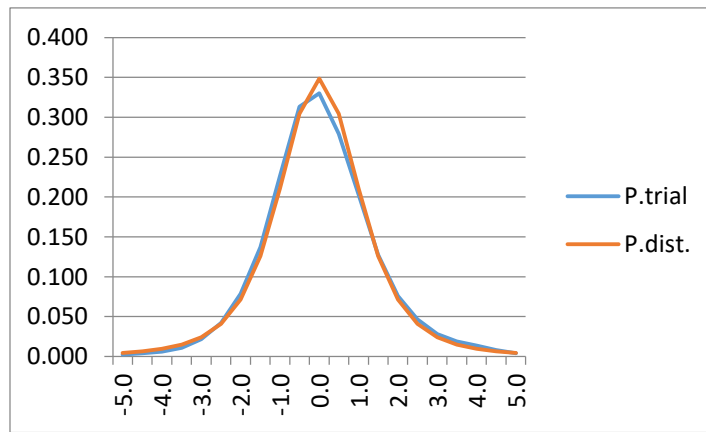
Q2: =T.DIST(O2,3,1)-T.DIST(N2,3,1)

R: =T.DIST(M2,3,1)

エクセル関数 T.DIST(t, df, s)は t (t 値), df (自由度), s (関数形式)という引数から、累積確率(s=1 の場合)を返します。自由度は標本サイズ(n) - 1 を使います。

この N2:O2 を t=5.0 の行までコピーすると、上の表が出来上がります。

ここで、乱数実験から得られた確率分布(P 列)と、エクセル関数による t 分布確率(Q 列)が近似する様子を次のグラフで確認します。



9.4.2. t 分布と正規分布

このようにして、t 値が自由度 n-1 の確率分布に従うことが実験で、確かめることができました。この標本のサイズ(n)が大きくなると、t 値の確率分布は、正規分布に近づきます。n=30 であれば、ほとんど正規分布と同じになることを次の実験で確かめます。

T	U	V	W	X	Y	Z	AA	AB	A
t	from	to	t.dist.1	t.dist.5	t.dist.10	t.dist.20	t.dist.30	Norm.dist.	
-5.0	-5.5	-4.5	0.012	0.002	0.000	0.000	0.000	0.000	
-4.5	-5.0	-4.0	0.015	0.003	0.001	0.000	0.000	0.000	
-4.0	-4.5	-3.5	0.019	0.005	0.002	0.001	0.001	0.000	
-3.5	-4.0	-3.0	0.024	0.010	0.005	0.003	0.003	0.001	
-3.0	-3.5	-2.5	0.033	0.019	0.013	0.009	0.008	0.006	
-2.5	-3.0	-2.0	0.045	0.036	0.030	0.026	0.025	0.021	
-2.0	-2.5	-1.5	0.066	0.070	0.067	0.064	0.063	0.061	
-1.5	-2.0	-1.0	0.102	0.131	0.134	0.135	0.135	0.136	
-1.0	-1.5	-0.5	0.165	0.222	0.232	0.237	0.238	0.242	
-0.5	-1.0	0.0	0.250	0.318	0.330	0.335	0.337	0.341	
0.0	-0.5	0.5	0.295	0.362	0.372	0.377	0.379	0.383	
0.5	0.0	1.0	0.250	0.318	0.330	0.335	0.337	0.341	
1.0	0.5	1.5	0.165	0.222	0.232	0.237	0.238	0.242	
1.5	1.0	2.0	0.102	0.131	0.134	0.135	0.135	0.136	
2.0	1.5	2.5	0.066	0.070	0.067	0.064	0.063	0.061	
2.5	2.0	3.0	0.045	0.036	0.030	0.026	0.025	0.021	
3.0	2.5	3.5	0.033	0.019	0.013	0.009	0.008	0.006	
3.5	3.0	4.0	0.024	0.010	0.005	0.003	0.003	0.001	
4.0	3.5	4.5	0.019	0.005	0.002	0.001	0.001	0.000	
4.5	4.0	5.0	0.015	0.003	0.001	0.000	0.000	0.000	
5.0	4.5	5.5	0.012	0.002	0.000	0.000	0.000	0.000	

$$W2: =T.DIST(\$V2,1,1)-T.DIST(\$U2,1,1)$$

$$X2: =T.DIST(\$V2,5,1)-T.DIST(\$U2,5,1)$$

$$Y2: =T.DIST(\$V2,10,1)-T.DIST(\$U2,10,1)$$

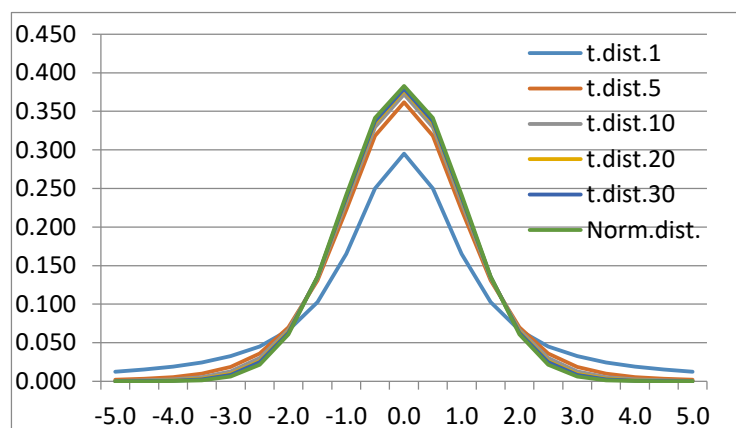
Z2: =T.DIST(\$V2,20,1)-T.DIST(\$U2,20,1)

AA2: =T.DIST(\$V2,30,1)-T.DIST(\$U2,30,1)

AB2: = =NORMSDIST(O2)-NORMSDIST(N2)

こうして設定した 2 行目を全体にコピーします。

次のグラフによって $n = 1, 5, 10, 20, 30$ の t 分布と正規分布のそれぞれの範囲の確率分布を比較すると、標本のサイズが大きくなるにつれて、標準正規分布に近づくことがわかります。



よって、標本のサイズが 30 以上であれば、正規分布による Z 検定が有効であると言えるでしょう。標本のサイズが 30 未満であれば、 t 検定を実行すべきです。

9.4.3. 平均の t 検定の例

次は従来の外国語教育法によるクラスの最終試験の平均点 $m (=60)$ と新しい方法によるクラスの最終試験の平均点 $M (=65)$ を比較して、平均点が有意に上昇したかどうかを調べるための t 検定の結果です。

t-test.0 (M:m).Result.SD	
m	60.000
N	10
M	65.000
u.V	100.000
u.SD	10.000
Significance level	0.050
Standard error	3.162
t-value	1.581
Lower probability	0.926
Lower critical value	-1.833
Lower significance	NS (H0)

Upper probability	0.074
Upper critical value	1.833
Upper significance	NS (H0)
Two-tailed probability	0.148
Two-tailed critical value	2.262
Two-tailed significance	NS (H0)

ここで次の式を用いています。

$$t\text{-value} := (M - m) / \text{SQRT}(VK5/K3)$$

$$\text{Lower probability} := \text{T.DIST}(t\text{-value}, N-1, 1)$$

$$\text{Upper probability} := 1 - \text{Lower probability}$$

このデータ (n=10, M = 65, u.SD=10) では、有意な結果になりませんが、点数(M)が 67 点であれば、危険率は 2.8% となって有意になります。また、65 点であっても、標本のサイズが 20 であれば、やはり危険率は 1.9% で有意になります。

9.5. 対応データの差平均の t 検定

9.5.1. 差平均の t 値の確率分布

次のようなデータ x, y の差(x-y)の平均 M(x-y)の t 値は、t(n-1)の確率分布を示します。

$$t = M(x-y) / [uV(x-y) / n]^{1/2}$$

n	x	y	x-y
1	26	15	11
2	19	16	3
3	21	14	7
4	16	15	1
5	25	20	5
6	17	15	2

このことを Excel で実験して確かめます。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	t-test.1			N	x1	x2	x3	x4	y1	y2	y3	y4	x1-y1	x2-y2	x3-y3	x4-y4	M(x-y)	u.V(x-y)	t
2	Set			1	8.6	8.2	8.0	7.5	7.1	10.9	7.3	12.4	1.5	-2.7	0.7	-5.0	-1.3	9.120	-0.892
3	mr	10.000		2	10.9	12.5	12.1	14.4	12.6	9.9	14.1	12.0	-1.7	2.6	-2.0	2.5	0.3	6.473	0.269
4	vr	9.000		3	10.3	12.5	11.1	2.3	9.8	8.3	8.6	9.3	0.5	4.2	2.5	-7.0	0.1	24.464	0.029
5	nx		4	4	13.2	7.0	13.7	11.1	6.0	11.2	6.6	13.6	7.2	-4.2	7.2	-2.4	1.9	37.247	0.632
6	ny		4	5	9.1	10.1	10.2	9.9	12.6	8.3	9.1	7.5	-3.5	1.8	1.1	2.4	0.5	7.133	0.345
7				6	6.4	8.4	4.4	5.6	11.8	8.8	7.3	9.2	-5.4	-0.3	-3.0	-3.5	-3.1	4.391	-2.920
8	Trial			7	9.8	8.6	8.3	7.3	10.9	15.0	11.0	6.0	-1.1	-6.4	-2.7	1.3	-2.2	10.581	-1.367
9	mx	9.918		8	11.4	9.3	13.1	1.8	10.4	13.8	6.9	10.8	1.0	-4.5	6.2	-9.1	-1.6	43.979	-0.475
10	my	10.004		9	6.8	8.9	10.1	10.0	12.3	8.5	4.5	7.7	-5.4	0.4	5.6	2.3	0.7	21.339	0.301
11	vx	9.015		10	10.1	5.4	12.6	7.9	11.5	8.5	7.5	12.5	-1.4	-3.1	5.1	-4.6	-1.0	18.594	-0.468
12	vy	8.605		11	8.4	9.9	11.8	11.2	18.0	9.0	13.5	1.0	-9.7	0.9	-1.7	10.2	-0.1	66.850	-0.021
13				12	11.3	5.2	11.8	10.7	11.3	8.6	12.4	10.4	0.0	-3.4	-0.7	0.3	-0.9	2.836	-1.120
14	Renew: [F9]			13	7.4	9.3	13.6	14.0	13.8	14.8	11.3	15.2	-6.4	-5.5	2.3	-1.2	-2.7	16.197	-1.349
15				14	10.6	0.4	12.8	0.2	10.2	4.0	5.8	0.6	0.4	4.2	7.4	0.2	2.0	11.052	1.650

E2: =NORMINV(RAND(),\$B\$3,SQRT(\$B\$4)) ←正規分布乱数

M2: =E2-I2 ←対応する値の差

Q2: =AVERAGE(M2:P2) ←差の平均

R2: =VAR(M2:P2) ←差の不偏分散

S2: =Q2/SQRT(R2/4) ←差の平均の t 値

上で求めた t 値(S 列)の実験で得られた確率分布(P.trial)と, 対応する t 値の理論的な確率分布(P.dist.)を比較します。

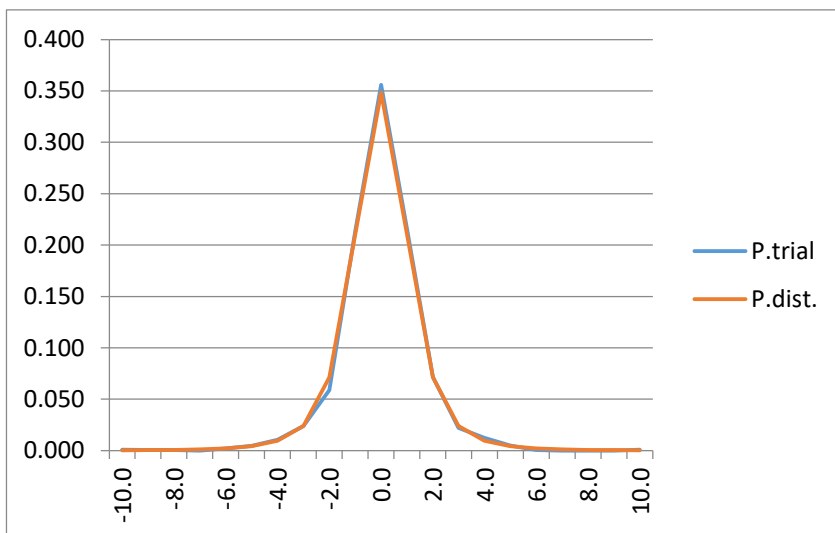
U	V	W	X	Y	Z	A
t	from	to	P.trial	P.dist.	P.cum.	
-10.0	-10.5	-9.5	0.001	0.000	0.001	
-9.0	-9.5	-8.5	0.000	0.000	0.001	
-8.0	-8.5	-7.5	0.001	0.001	0.002	
-7.0	-7.5	-6.5	0.001	0.001	0.003	
-6.0	-6.5	-5.5	0.003	0.002	0.005	
-5.0	-5.5	-4.5	0.002	0.004	0.008	
-4.0	-4.5	-3.5	0.012	0.009	0.014	
-3.0	-3.5	-2.5	0.018	0.024	0.029	
-2.0	-2.5	-1.5	0.074	0.071	0.070	
-1.0	-1.5	-0.5	0.192	0.210	0.196	
0.0	-0.5	0.5	0.354	0.349	0.500	
1.0	0.5	1.5	0.220	0.210	0.804	
2.0	1.5	2.5	0.069	0.071	0.930	
3.0	2.5	3.5	0.034	0.024	0.971	
4.0	3.5	4.5	0.009	0.009	0.986	
5.0	4.5	5.5	0.005	0.004	0.992	
6.0	5.5	6.5	0.003	0.002	0.995	
7.0	6.5	7.5	0.001	0.001	0.997	
8.0	7.5	8.5	0.001	0.001	0.998	
9.0	8.5	9.5	0.001	0.000	0.999	
10.0	9.5	10.5	0.001	0.000	0.999	

X2: =COUNTIFS(\$S\$2:\$S\$2001,">"&V2,\$S\$2:\$S\$2001,"<="&W2)/2000

Y2: =T.DIST(W2,3,1)-T.DIST(V2,3,1)

Z2: =T.DIST(U2,3,1)

下図を見ると、P.trial と P.dist が強く近似していることがわかります。



9.5.2. 差平均の t 検定

差の平均が自由度 $n-1$ の t 分布になる、という性質を利用して、差の平均が有意であるか否かを調べること（「対応のある t 検定」）ができます。

n	x	y	x-y
1	26	15	11
2	19	16	3
3	21	14	7
4	16	15	1
5	25	20	5
6	17	15	2

t-test.1 (M1-M2) paired test	
M(X-Y)	4.8333
N	6.0000
u.V(X-Y)	13.7667
u.SD(X-Y)	3.7103
Significance level	0.0500
t-value	3.1909
Lower probability	0.9879
Lower critical value	-2.0150
Lower significance	NS (H0)
Upper probability	0.0121
Upper critical value	2.0150
Upper significance	S. (H1)
Two-tailed probability	0.0242
Two-tailed critical value	2.5706
Two-tailed significance	S. (H1)

$$t\text{-value} = G2/\text{SQRT}(G4/G3) = M(X-Y) / [u.V(X-Y) / N]^{1/2}$$

$$\text{Lower probability} = \text{T.DIST}(t\text{-value}, N-1, 1)$$

Upper probability = 1 - Lower probability

Upper significance

=IF(Upper probability < Significance level,"S. (H1)", "NS (H0)")

次は Excel の分析ツールにある「t-検定: 一对の標本による平均の検定ツール」の結果です。この結果は上の分析と一致します。

	x	y
平均	20.6667	15.8333
分散	17.0667	4.5667
観測数	6	6
ピアソン相関	0.4455	
仮説平均との差異	0	
自由度	5	
t	3.1909	
P(T<=t) 片側	0.0121	
t 境界値 片側	2.0150	
P(T<=t) 両側	0.0242	
t 境界値 両側	2.5706	

9.6. 非対応等分散データの平均差 t 検定

9.6.1. 平均差の t 値の確率分布

2つのデータセットの平均の差(M1 - M2)をその「標準誤差」(se: standard error)で割った値の確率の分布は、自由度(n1 + n2 - 2)の t 分布確率になることが知られています。これを Excel で実験して確かめます。ここで、「標準誤差」(se)は次の式で計算されます。

$$se = [uv1 (n1 - 1) + uv2 (n2 - 1) / (n1 + n2 - 2)]^{1/2} * (1/n1 + 1/n2)^{1/2}$$

ここで uv1, uv2 はそれぞれのデータの不偏分散であり、n1, n2 はそれぞれのデータのサイズです。

よって、t 値は

$$t = (m1 - m2) / se$$

ここで m1, m2 はそれぞれのデータの平均です。

次のシートでこのようにして定義した t 値の確率分布を求めます。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	t.3 dist	(same var.)		Trial	x1	x2	x3	x4	y1	y2	y3	Mx	My	Vx	Vy	t
2	Set			1	10.3	10.0	7.6	11.2	10.5	9.1	8.1	9.8	9.2	2.440	1.485	0.481
3	mr	10.000		2	8.0	11.3	8.9	10.4	10.9	10.0	11.2	9.6	10.7	2.260	0.355	-1.109
4	vr	2.000		3	9.7	11.8	8.8	10.6	10.9	12.1	12.6	10.2	11.9	1.681	0.826	-1.897
5	nx	4		4	11.7	10.0	11.0	10.5	12.6	9.6	11.6	10.8	11.3	0.488	2.362	-0.573
6	ny	3		5	10.6	9.5	9.7	8.8	11.3	8.5	8.5	9.6	9.4	0.568	2.538	0.264
7				6	10.2	10.8	8.2	13.1	10.3	10.5	8.4	10.6	9.7	3.995	1.439	0.618
8	Trial			7	4.8	9.3	10.2	11.2	9.1	9.3	9.1	8.9	9.2	7.819	0.015	-0.178
9	mx	9.956		8	8.8	7.2	8.7	9.0	11.0	12.6	12.6	8.4	12.1	0.635	0.880	-5.607
10	my	9.981		9	8.6	8.6	10.0	9.6	10.1	7.9	10.9	9.2	9.6	0.494	2.514	-0.465
11	vx	1.993		10	10.2	8.3	10.9	11.3	11.3	6.6	8.7	10.2	8.8	1.839	5.561	0.972
12	vy	1.959		11	7.8	8.4	9.1	9.7	8.8	9.8	10.2	8.8	9.6	0.685	0.502	-1.472
13				12	12.8	9.0	8.5	12.1	8.8	9.0	9.4	10.6	9.0	4.699	0.081	1.206
14	Renew: [F9]			13	10.5	8.4	10.6	11.1	7.8	10.3	10.1	10.2	9.4	1.397	1.870	0.799
15				14	7.7	12.1	10.1	10.3	6.9	11.8	8.6	10.0	9.1	2.361	6.166	0.505

E2: =NORMINV(RAND(),\$B\$3,SQRT(\$B\$4)) ← x1

L2: =AVERAGE(E2:H2) ← Mx

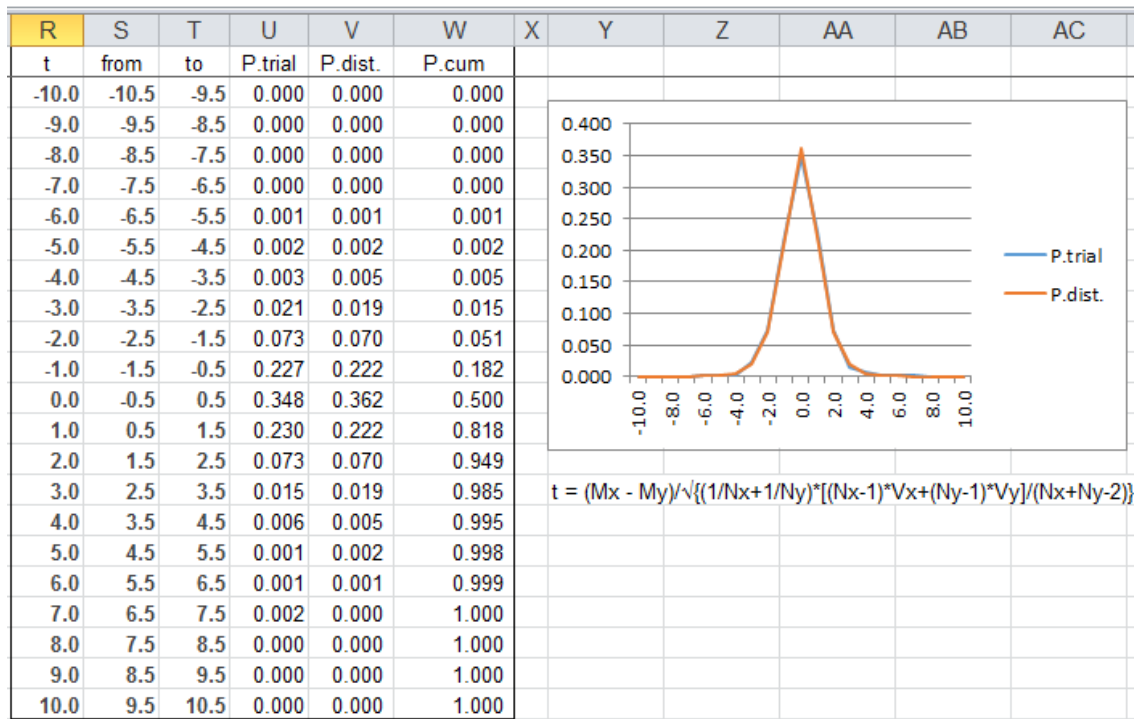
M2: =AVERAGE(I2:K2) ← My

N2: =VAR(E2:H2) ← Vx

O2: =VAR(I2:K2) ← Vy

P2: =(L2-M2)/SQRT((1/4+1/3)*((3*N2+2*O2)/5)) ← t

次に実験で求めた t 値の確率分布と，理論的な t 値の確率分布の確率分布が一致することを確認めます。



U2: =COUNTIFS(\$P\$2:\$P\$2001,">"&S2,\$P\$2:\$P\$2001,"<="&T2)/2000 ← P.trial

V2: =T.DIST(T2,5,1)-T.DIST(S2,5,1) ← P.dist

W2: =T.DIST(R2,5,1) ← P.cum.

9.6.2. 平均差の t 検定

先に確かめた t 値の確率分布を使った差の t 検定を行います。データ例は次のシートの A-C 列です。X と Y の平均の差が有意であるか、否かを検定します。

	A	B	C	D	E	F	G	H	I	J	K
1	N	X	Y		t.3.test (same var.). Data	X	Y				
2	1	33	30		N	5.0000	7.0000				
3	2	37	31		M	34.6000	33.1429		t-検定: 等分散を仮定した 2 標本による検定		
4	3	35	32		V	4.3000	5.1429				
5	4	36	35		Significance level	0.0500				X	Y
6	5	32	36		Degree of freedom	10.0000			平均	34.6000	33.1429
7	6		35		Standard deviation	2.0736	2.2678		分散	4.3000	5.1429
8	7		33		Standard error	0.9274	0.8571		観測数	5.0000	7.0000
9					t-value	1.1352			プールされた分散	4.8057	
10					Lower probability	0.8586			仮説平均との差異	0.0000	
11					Lower critical value	-1.8125			自由度	10.0000	
12					Lower significance	NS (H0)			t	1.1352	
13					Upper probability	0.1414			P(T<=t) 片側	0.1414	
14					Upper critical value	1.8125			t 境界値 片側	1.8125	
15					Upper significance	NS (H0)			P(T<=t) 両側	0.2828	
16					Two-tailed probability	0.2828			t 境界値 両側	2.2281	
17					Two-tailed critical value	2.2281					
18					Two-tailed significance	NS (H0)					

t.3.test (same var.). Data X Y

N =COUNT(B:B) =COUNT(C:C)

M =AVERAGE(B:B) =AVERAGE(C:C)

V =VAR.S(B:B) =VAR.S(C:C)

Significance level 0.05

Degree of freedom =F2+G2-2

Standard deviation =SQRT(F4) =SQRT(G4)

Standard error =SQRT(F4/F2) =SQRT(G4/G2)

t-value =(F3-G3)/SQRT(((1/F2+1/G2)*(((F2-1)*F4+(G2-1)*G4)/(F6))))

Lower probability =T.DIST(F9,F6,1)

Lower critical value =T.INV(F5,F6)

Lower significance =IF(F10<F5,"S. (H1)", "NS (H0)")

Upper probability =1-F10

Upper critical value =T.INV(1-F5,F6)

Upper significance =IF(F13<F5,"S. (H1)", "NS (H0)")

Two-tailed probability =TDIST(ABS(F9),F6,2)

Two-tailed critical value =TINV(F5,F6)

Two-tailed significance =IF(F16<F5,"S. (H1)", "NS (H0)")

このデータ例では P = 0.141 なので有意とは言えず、帰無仮説を棄却できません。右の出力は Excel の「データ分析」の結果です。同じ値を返しています。

9.7. 非対応異分散データの平均差 t 検定

9.7.1. 平均差の t 値の確率分布

差の検定の対象となる 2 つのデータセットが、等分散とはみなせない場合には Welch 法が使われます¹²。そのときの t 値は次のように定義されます。

$$t = (m1 - m2) / (uv1/n1 + uv2/n2)^{1/2}$$

ここで、m1, m2 はそれぞれのデータの平均値、uv1, uv2 はそれぞれの不偏分散、n1, n2 はそれぞれのデータ数を示します。

この t 値の確率分布は自由度(df)が次の式で定義された t 分布になることが知られています。

$$df = (uv1/n1 + uv2/n2)^2 / [(uv1/n1)^2/(n1-1) + (uv2/n2)^2/(n1-1)]$$

この複雑な式が返す値の小数点以下を四捨五入し、その整数部(Int)を自由度とします¹³。次は、正規乱数で発生させた 2 つのデータセットの t 値を 2000 回計算したときのシートの最初の部分です。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	t.4.dist	(diff. var.)	Trial	x1	x2	x3	x4	y1	y2	y3	m.x	m.y	uv.x	uv.y	t		
2	Set		1	12.4	9.6	15.8	1.7	10.0	10.2	12.1	9.9	10.8	36.2	1.4	-0.290		
3	m	10.000	2	7.1	0.4	7.3	11.3	12.8	6.2	5.3	6.5	8.1	20.502	16.8	-0.470		
4	sdx	5.000	3	-0.6	16.0	17.1	10.1	14.9	10.5	16.6	10.7	14.0	65.963	10.0	-0.745		
5	sdv	3.000	4	6.5	4.5	11.8	15.2	7.6	12.0	6.9	9.5	8.8	24.083	7.6	0.232		
6	nx	4	5	15.8	11.2	6.8	10.0	11.8	14.7	11.1	11.0	12.6	14.068	3.7	-0.730		
7	ny	3	6	15.9	10.2	11.9	9.3	13.9	6.8	9.6	11.8	10.1	8.573	12.8	0.684		
8			7	9.9	10.2	8.9	15.3	10.5	10.4	8.7	11.1	9.9	8.261	0.9	0.804		
9	Trial		8	20.3	0.9	8.7	1.5	9.7	13.8	5.5	7.8	9.7	81.294	17.5	-0.357		
10	Mx	10.045	9	4.3	9.9	6.4	15.5	6.8	10.3	10.3	9.0	9.1	24.364	4.0	-0.042		
11	My	9.941	10	9.4	1.9	15.1	1.6	6.4	12.6	16.0	7.0	11.7	42.059	23.5	-1.088		
12	SDx	5.039	11	3.7	10.3	19.4	14.6	7.5	4.3	10.4	12.0	7.4	44.580	9.3	1.215		
13	SDy	3.018	12	-0.2	11.5	6.8	5.5	7.0	9.8	10.7	5.9	9.2	23.271	3.7	-1.229		
14			13	4.2	9.0	5.2	6.4	10.6	9.2	3.2	6.2	7.7	4.381	15.4	-0.590		
15	df	5.000	14	14.3	11.5	8.0	12.1	10.6	12.7	13.7	11.5	12.3	6.857	2.5	-0.524		

Mx: B10: =AVERAGE(E:H)

My: B11: =AVERAGE(I:K)

SDx: B12: =STDEVP(E:H)

SDy: B13: =STDEVP(I:K)

¹² おおよその目安は分散やデータ数が 2 倍以上違うときに使われますが、分散が異なることを精確に調べるときは、F 検定を使います。→分散の F 検定

¹³ 一般に自由度が小さい方が、検定の確率値が大きくなるので、厳しく検定するために、小数点以下を切り捨てる方法がとられていますが、逆に厳し過ぎる検定であると、有意差を見失う可能性もあります。一方、小数点以下を切り上げると、甘すぎる検定になります。そこで Excel では、自由度として、四捨五入した整数値を使用しているようです。

df: B15: =ROUND((B12^2/4+B13^2/3)^2/(B12^2/4)^2/3+(B13^2/3)^2/2,0)

m.x: L2: =AVERAGE(E2:H2)

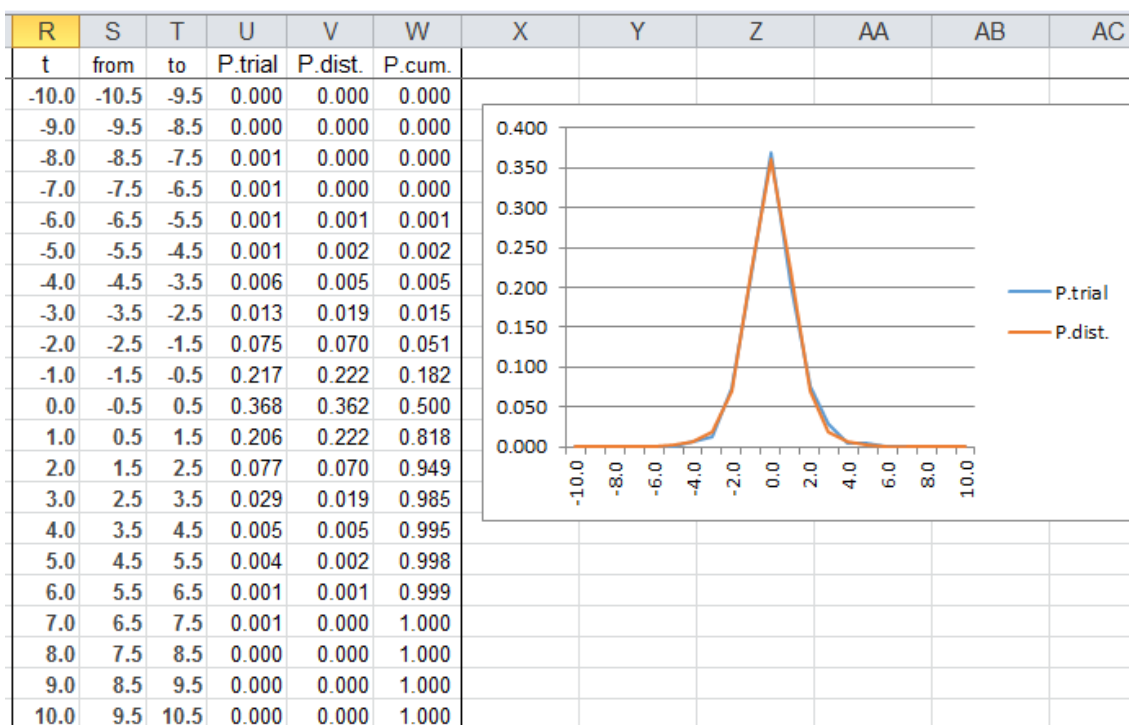
m.y: M2: =AVERAGE(I2:K2)

uv.x: N2: =VAR(E2:H2)

uv.y: O2: =VAR(I2:K2)

t: P2: =(L2-M2)/SQRT(N2/4+O2/3)

次に、実験で得られた t 値の確率分布と、理論的な t 分布の確率を比較し、両者が近似することを観察します。



U2: =COUNTIFS(\$P\$2:\$P\$2001,">"&S2,\$P\$2:\$P\$2001,"<="&T2)/2000

V2: =T.DIST(T2,\$B\$15,1)-T.DIST(S2,\$B\$15,1)

W2: =T.DIST(R2,\$B\$15,1)

9.7.2. 平均差の t 検定

Welch の方法を使って、平均差の t 検定を行います。

	A	B	C	D	E	F	G	H	I	J	K	L
1	N	X	Y		t-test.3 (diff. var.) Data	X	Y					
2	1	33	30		N	5.0000	8.0000		t検定: 分散が等しくないと仮定した2標本による検定			
3	2	31	31		M	33.2000	35.1250					
4	3	35	32		V	5.2000	21.2679					
5	4	36	39		Significance level	0.0500			平均	33.2000	35.1250	
6	5	31	44		Degree of freedom	11.0000			分散	5.2000	21.2679	
7	6		34		Standard deviation	2.2804	4.6117		観測数	5.0000	8.0000	
8	7		36		Standard error	1.0198	1.6305		仮説平均との差異	0.0000		
9	8		35		t-value	-1.0010			自由度	11.0000		
10					Lower probability	0.1692			t	-1.0010		
11					Lower critical value	-1.7959			P(T<=t) 片側	0.1692		
12					Lower significance	NS (H0)			t境界値 片側	1.7959		
13					Upper probability	0.8308			P(T<=t) 両側	0.3384		
14					Upper critical value	1.7959			t境界値 両側	2.2010		
15					Upper significance	NS (H0)						
16					Two-tailed probability	0.3384						
17					Two-tailed critical value	2.2010						
18					Two-tailed significance	NS (H0)						

t-test.3 (diff. var.) Data X Y

N =COUNT(B:B) =COUNT(C:C)

M =AVERAGE(B:B) =AVERAGE(C:C)

V =VAR.S(B:B) =VAR.S(C:C)

Significance level 0.05

Degree of freedom

=ROUND((F4/F2+G4/G2)^2/((F4/F2)^2/(F2-1)+(G4/G2)^2/(G2-1)),0)

Standard deviation =SQRT(F4) =SQRT(G4)

Standard error =SQRT(F4/F2) =SQRT(G4/G2)

t-value =(F3-G3)/SQRT(F4/F2+G4/G2)

Lower probability =T.DIST(F9,F6,1)

Lower critical value =T.INV(F5,F6)

Lower significance =IF(F10<F5,"S. (H1)", "NS (H0)")

Upper probability =1-F10

Upper critical value =T.INV(1-F5,F6)

Upper significance =IF(F13<F5,"S. (H1)", "NS (H0)")

Two-tailed probability =TDIST(ABS(F9),F6,2)

Two-tailed critical value =TINV(F5,F6)

Two-tailed significance =IF(F16<F5/2,"S. (H1)", "NS (H0)")

ここでは、Xの平均よりもYの平均が有意に大きいかどうかを検定します。逆には、Xの平均がYの平均より有意に小さいかどうか、という問題になります。そこで、下側確率(Lower probability)を見ると、16.9%を示しているため、帰無仮説を棄却できません。よって、有意な差ではないことがわかります。Excelの「t検定: 分散が等しくないと仮定した2標本による検定」も同じ結果を示しています。

9.8. 分散の F 検定

9.8.1. 分散比 F の確率分布

F 検定は 2 つのデータセットの分散が有意に異なるか・否かを検査する方法です。はじめに、次のように定義された分散比(F 値)の確率分布が F 分布に従うことを、Excel を使って実験して確かめます。

$$F = uv1 / uv2$$

ここで uv1, uv2 はそれぞれのデータの不偏分散を示します。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	F-test			Trial	x1	x2	x3	x4	y1	y2	y3	y4	y5	F
2	Set			1	17.1	17.4	22.2	23.4	20.1	18.5	25.0	23.2	21.4	1.624
3	m.r	20.000		2	16.7	18.6	22.5	23.2	17.6	18.3	22.8	17.2	17.3	1.744
4	v.r	9.000		3	15.5	19.3	18.8	17.7	15.1	18.9	17.5	20.9	21.4	0.424
5	nx	4		4	22.0	20.2	21.4	24.1	21.0	20.6	20.3	20.9	19.3	5.612
6	ny	5		5	17.9	20.9	19.4	17.9	16.7	24.3	20.4	19.9	19.2	0.271
7				6	20.6	17.4	19.3	13.6	19.0	18.0	20.0	12.3	20.0	0.913
8	Result:			7	20.1	21.5	15.2	19.7	13.4	20.3	18.4	18.3	23.1	0.598
9	m	19.985		8	16.1	14.0	20.5	21.1	24.2	17.2	21.6	25.2	21.6	1.210
10	v	8.998		9	23.4	21.2	18.9	19.3	20.4	25.3	21.9	20.7	21.6	1.114
11				10	23.3	18.3	21.9	16.9	21.8	20.5	23.1	13.4	17.9	0.615
12	Renew: [F9]			11	18.1	27.8	11.6	28.1	15.5	14.4	23.5	21.9	16.1	3.770
13				12	21.3	20.8	19.3	14.3	22.3	19.5	18.9	17.9	22.0	2.726

x1: E1: =NORMINV(RAND(), \$B\$3, SQRT(\$B\$4))

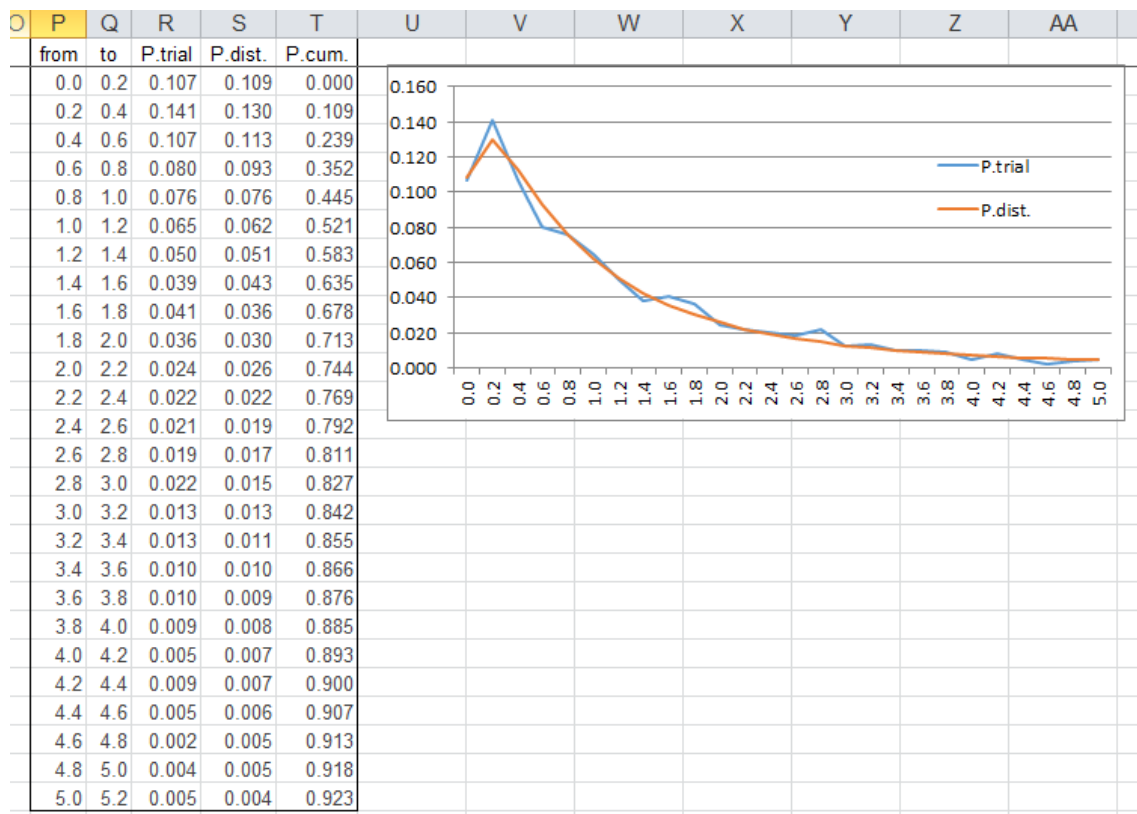
y1: I1: =NORMINV(RAND(), \$B\$3, SQRT(\$B\$4))

F: N1: =VAR(E2:H2)/VAR(I2:M2)

m: B9: =AVERAGE(E:M)

v: B10: =VAR.P(E:M)

このようにして実験で求めた分散比(F)は、次のグラフが示すように、理論的な F の確率分布と近似します。



R2: =COUNTIFS(\$N\$2:\$N\$2001,">"&P2,\$N\$2:\$N\$2001,"<"&Q2)/2000

S2: =F.DIST(Q2,3,4,1)-F.DIST(P2,3,4,1)

T2: =F.DIST(P2,3,4,1)

9.8.2. 分散の検定

次のシート上で、F 検定によって先に見た分散比が有意に大きな値であるか否かを確認めます。

	A	B	C	D	E	F	G	H	I	J	K
1	N	X	Y		F-test. Data	X	Y		F-検定: 2 標本を使った分散の検定		
2	1	33	30		N	5.0000	8.0000				
3	2	31	31		M	33.2000	35.1250			X	Y
4	3	35	32		V	5.2000	21.2679		平均	33.2000	35.1250
5	4	36	39		Significance level	0.0500			分散	5.2000	21.2679
6	5	31	44		Degree of freedom	4.0000	7.0000		観測数	5.0000	8.0000
7	6		34		F-value	0.2445			自由度	4.0000	7.0000
8	7		36		Upper probability	0.0958			観測された分散比	0.2445	
9	8		35		Upper critical value	0.1641			P(F<=f) 片側	0.0958	
10	9				Upper significance	NS (H0)			F 境界値 片側	0.1641	
11											

E	F	G
F-test. Data	X	Y
N	=COUNT(B:B)	=COUNT(C:C)
M	=AVERAGE(B:B)	=AVERAGE(C:C)
V	=VAR.S(B:B)	=VAR.S(C:C)
Significance level	0.05	
Degree of freedom	=F2-1	=G2-1
F-value	=F4/G4	
Upper probability	=IF(F4>G4, FDIST(I	
Upper critical value	=F.INV(F5,F6,G6)	
Upper significance	=IF(F8<F5,"S. (H1)'	

上のシート上の計算と Excel 分析ツール「F-検定: 2 標本を使った分散の検定」の結果は同じで、どちらも、有意な確率を返していません。よって、分散は等しい、という帰無仮説を棄却できません。

F 検定の結果は、非対応等分散の平均差の検定と、非対応異分散の平均差の検定のどちらを選択するかを決定するときに使われます。

9.9. カイ二乗検定

次のような表から連関度を数値化する方法を扱います¹⁴。

語	and	but	so	合計
全体	58	43	28	129

単純集計表

語	and	but	so	合計
文頭	12	7	11	30
文中	46	36	17	99
全体	58	43	28	129

クロス集計表

上の表は 1 つの指標（英語の等位接続詞）について数値（頻度）を表したものです。一方、下の表は(1)「英語の等位接続詞」と(2)「出現位置」という 2 つの指標を基に集計したものです。このような表は「クロス集計表」とよばれます。ここで問題となるのは、この 2 つの指標はお互いに関連しているかどうかということです。具体的に言うと、2 つが関係しているとき「(1)英語の等位接続詞の(2)出現位置は単語によって異なる」という結論になりますし、関係していないとき「(1)英語の等位接続詞の(2)出現位置は単語に左右されない」（それぞれの現象は「独立」である）ということになります。この判定をする手法が「カイ二乗検定」(chi square test)です。

¹⁴ * 参考：池田央．1976.『統計的方法 I 基礎』新曜社． pp.121-132.

ここではカイ二乗検定を理解するために簡単な例として 2 x 2 の表を用いて説明します。

次のようなケースを考えてみましょう。ある現象を数えるにあたって、次のように、それが出現した場合だけを数えるやり方があります。

「教育方法 A」に効果があったケース：59

「教育方法 B」に効果があったケース：49

「方法 A」に効果があった場合の数を 59、「方法 B」に効果があった場合の数を 49 として単純に比較すると、確かに「方法 A」のほうが優れている、という結論になるかもしれませんが、しかし、ここで「方法 A」（そして「方法 B」）に効果があったことを確かめるには、「方法 A」（そして「方法 B」）に効果がなかったケースも調べる必要があります。その結果が次の表です。

実測値1	効果がある	効果がない
方法A	59	35
方法B	49	53

実測値 1 の結果を見ると、やはり「方法 A」のほうが「方法 B」より優れているように見えますが、仮に次の実測値 2 ようなケースになったときは判断が逆転してしまいます。

実測値2	効果がある	効果がない
方法A	59	65
方法B	49	55

「方法 A」と「方法 B」はどちらも効果がある場合よりも効果がない場合の方が上回り、それぞれの方法の差は 6 ですが、「方法 A」の「効果がない」の数が大きくなっています。

さらに、次の実測値 3 のようなケースがあります。「方法 A」も「方法 B」もどちらも「効果がある」の数が「効果がない」の数よりも上回っています。両者は「効果がある」と「効果がない」の差は 10 となっています。はたして「方法 A」が「方法 B」に比べて効果があると言えるのでしょうか。

実測値3	効果がある	効果がない
方法A	59	49
方法B	49	39

これらは単に「効果がある」という肯定的な反応だけを数えていては見つからなかった問題を示しています。つまり、方法 A と方法 B の差を考えるには、効果があった場合と効果がなかった場合の両方を考える必要があるということです。

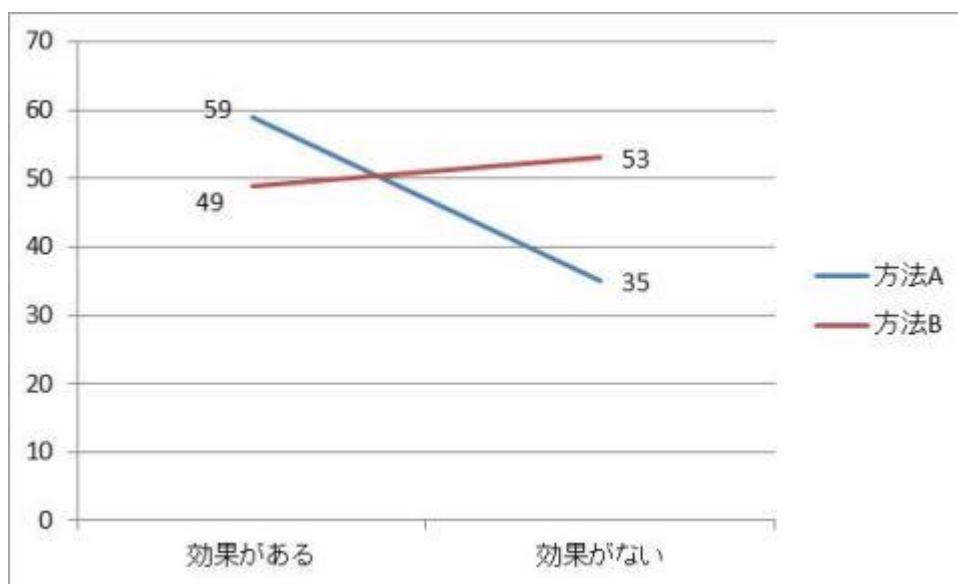
それでは実測値 1~3 の場合、方法 A と方法 B に差があるといえるのはどれでしょうか。この数値を統計的に算出するのが「カイ二乗検定」(chi-square test)です。この方法を用いることで、差があるかどうかをはっきりと数値で示すことができます。

9.9.1. 期待値

カイ二乗検定の基になるデータは、(1)実測値と(2)期待値です。以下、例として実測値 1 を見ていきましょう¹⁵。

実測値1	効果がある	効果がない	計	比率
方法A	59	35	94	1.686
方法B	49	53	102	0.925
計	108	88	196	

このデータをグラフにして視覚化しておきます。



説明のために、観測値を次のようによぶことにします。

実測値	効果がある	効果がない
方法A	a	b
方法B	c	d

次の表では各セルに「期待される」得点(期待得点：expected score)が示されています。期待得点は次のような式で計算されます。

¹⁵ 比率は「効果がある」の値を「効果がない」の値で割った値です。1 より大きいと「効果がある」ほうが多いことを示します。

期待値	効果がある	効果がない
方法 A	$\frac{(a+b) \times (a+c)}{a+b+c+d}$	$\frac{(a+b) \times (b+d)}{a+b+c+d}$
方法 B	$\frac{(c+d) \times (a+c)}{a+b+c+d}$	$\frac{(c+d) \times (b+d)}{a+b+c+d}$

たとえば、方法 A の効果がある期待値は、方法 A の総数(a + b) 94 に「効果がある」と期待できる率(a + c) 108/総数(a+b+c+d)を掛けた値です。総数 94 が 108:88 に分割されるときに 108 の側に当然期待できる数値, $94 \times (108 / 196)$ を示します¹⁶。

- 「方法 A」・「効果がある」の期待得点 $94 * 108 / 196 = 51.796$
- 「方法 A」・「効果がない」の期待得点 $94 * 88 / 196 = 42.204$
- 「方法 B」・「効果がある」の期待得点 $102 * 108 / 196 = 56.204$
- 「方法 B」・「効果がない」の期待得点 $102 * 88 / 196 = 45.796$

期待値	効果がある	効果がない	計
方法A	51.796	42.204	94
方法B	56.204	45.796	102
計	108	88	196

もし実際に観察される値が当然予測される値（期待値）と近いならば、「偶然でも起こるかもしれない分布」ということになります。逆に、もしそれが期待値から大きく外れるならば、観察されたデータは有意な分布を示していると考えられます。「偶然ではほとんどあり得ない」と考えるのです。つまり、カイ二乗検定のポイントは「実測値と期待値のズレを見る」ことです。

9.9.2. カイ二乗値

期待値と実測値のずれを総合的に判断するため、すべての升目(a, b, c, d)における実際の観測値と期待値の「相対的な差」の総和で求めます。相対化するには、実測値から期待値を引いたものを期待値で割ります。また、「相対的な差」の合計は、そのままでは 0 になってしまうので、期待値からの距離を求めるために二乗しておきます（これがカイ「二乗」という名前の由来です）。

¹⁶ ここで、これらの期待値のすべてが 5 以上であるかどうかを確かめておきます。いずれかが 5 以下だと誤差が大きくてカイ二乗検定には向いていないデータと判断されます。

$$\text{標準化した値} = \frac{(\text{実測値} - \text{期待値})^2}{\text{期待値}}$$

このような操作を「標準化」とよびます。データには一定の単位がありますが、標準化すると単位がなくなります。単位がなくなると、どのようなデータでも統計的に同じ処理ができるようになるのです。次がそれぞれの相対的な差です。

標準化	効果がある	効果がない
方法A	1.002	1.230
方法B	0.923	1.133

これらの値を合計した値が「カイ二乗の統計量」 (χ^2) とよばれるものです。

$$\chi^2 = 1.002 + 1.230 + 0.923 + 1.133 = 4.288$$

式を一般化しましょう。実測値 a, b, c, d の期待値をそれぞれ, a', b', c', d' とし、標準化した値の総和は次のようになります。

$$\chi^2 = \frac{(a - a')^2}{a'} + \frac{(b - b')^2}{b'} + \frac{(c - c')^2}{c'} + \frac{(d - d')^2}{d'}$$

(上式の a', b', c', d' は a, b, c, d にそれぞれ対応する期待値を示します。) カイ二乗値は、期待値からのズレ (距離) の総和ということになります。この値が大きいほど、期待値とのズレが大きいといえることができます。

以上で 2*2 のクロス表について見ましたが、同じことは一般に次のような n*p のクロス表でも考えられます。

F	1	2	...	p	和
1	f.11	f.12	...	f.1p	s.1
2	f.21	f.22	...	f.2p	s.2
...
n	f.n1	f.n2	...	f.np	s.n
和	t.1	t.2	...	t.p	T

頻度を F = f.ij, 期待値を E = e.ij とすればカイ二乗値 (χ^2) は次の式で示されます。

$$\chi^2 = \sum_{i:1,n} \sum_{j:1,p} (f.ij - e.ij)^2 / e.ij$$

上の式の期待値 e.ij は

$$e.ij = s.i * t.j / T$$

よって

$$\chi^2 = \sum_{i:1,n} \sum_{j:1,p} (f_{ij} - s_{i \cdot} t_{\cdot j} / T)^2 / (s_{i \cdot} t_{\cdot j} / T)$$

●カイ二乗値の別式(1)

2*2 の表のカイ二乗値を求める式(1)は、簡単な式(2)に書き換えられます。

$$\chi^2 = \frac{(a - a')^2}{a'} + \frac{(b - b')^2}{b'} + \frac{(c - c')^2}{c'} + \frac{(d - d')^2}{d'} \quad \dots(1)$$

$$\chi^2 = N (ad - bc)^2 / (e f g h) \quad \dots(2)$$

ここで、N, a, b, c, d, e, f, g, h は次の表に示された値です。

X.Y	y1	y2	和
x1	a	b	e = a + b
x2	c	d	f = c + d
和	g = a + c	h = b + d	N = a + b + c + d

小寺(2002: 179-180)に従って、(1)の式から(2)の式を導きます。はじめに a に対応するカイ二乗値を求めます。

$$\begin{aligned} (a - a')^2 / a' &= (a - eg / N)^2 / (eg / N) \\ &= [a - (a+b)(a+c) / (a+b+c+d)]^2 / (eg / N) \\ &= \{ [a(a+b+c+d) - (a+b)(a+c)] / (a+b+c+d) \}^2 / (eg / N) \\ &= [(a^2 + ab + ac + ad) - (a^2 + ac + ab + bc)]^2 / (eg / N) \\ &= (ad - bc)^2 / (eg / N) \end{aligned}$$

同様にして、b, c, d に対応するカイ二乗値を求めます。

$$\begin{aligned} (b - b') / b' &= (ad - bc)^2 / (eh / N) \\ (c - c') / c' &= (ad - bc)^2 / (fg / N) \\ (d - d') / d' &= (ad - bc)^2 / (fh / N) \end{aligned}$$

よって表全体のカイ二乗値は

$$\begin{aligned} \chi^2 &= \frac{(a - a')^2}{a'} + \frac{(b - b')^2}{b'} + \frac{(c - c')^2}{c'} + \frac{(d - d')^2}{d'} \\ &= (ad - bc)^2 [1 / (eg) + 1 / (eh) + 1 / (fg) + 1 / (fh)] / N \\ &= (ad - bc)^2 [(fh + fg + eh + eg) / (e f g h)] / N \\ &= (ad - bc)^2 [(e + f)(g + h) / (e f g h)] / N \\ &= (ad - bc)^2 [N^2 / (efgh)] / N \quad \leftarrow \text{上表 : } e + f = N, g + h = N \\ &= N (ad - bc)^2 / (e f g h) \end{aligned}$$

参考：小寺平治 (2002) 『ゼロから学ぶ統計解析』講談社 p. 180.

● カイ二乗値の別式(2)

一般に $n * p$ の表のカイ二乗値を求める式は次の簡単な式に書き換えられます (伊藤 2002:109)。

$$\chi^2 = T [\sum_{i:1,n} \sum_{j:1,p} f_{.ij} / (s_i * t_j) - 1] \quad \leftarrow \text{簡単な別式}$$

以下で先に見たカイ二乗値の複雑な一般式

$$\chi^2 = \sum_{i:1,n} \sum_{j:1,p} (f_{.ij} - s_i * t_j / T)^2 / (s_i * t_j / T) \quad \leftarrow \text{一般式}$$

から上の簡単な別式を導きます。

その準備として n 行 p 列のクロス表を再掲します。

F	1	2	...	p	和
1	f.11	f.12	...	f.1p	s.1
2	f.21	f.22	...	f.2p	s.2
...
n	f.n1	f.n2	...	f.np	s.n
和	t.1	t.2	...	t.p	T

はじめに一般式の Σ を展開します。

$$\begin{aligned} \chi^2 &= \sum_{i:1,n} \sum_{j:1,p} (f_{.ij} - s_i * t_j / T)^2 / (s_i * t_j / T) \\ &= (f_{.11} - s_{.1} * t_{.1} / T)^2 / (s_{.1} * t_{.1} / T) \\ &+ (f_{.12} - s_{.1} * t_{.2} / T)^2 / (s_{.1} * t_{.2} / T) \\ &+ \dots \\ &+ (f_{.21} - s_{.2} * t_{.1} / T)^2 / (s_{.2} * t_{.1} / T) \\ &+ (f_{.22} - s_{.2} * t_{.2} / T)^2 / (s_{.2} * t_{.2} / T) \\ &+ \dots \\ &+ (f_{.n1} - s_{.n} * t_{.1} / T)^2 / (s_{.n} * t_{.1} / T) \\ &+ (f_{.n2} - s_{.n} * t_{.2} / T)^2 / (s_{.n} * t_{.2} / T) \\ &+ \dots \\ &+ (f_{.np} - s_{.n} * t_{.p} / T)^2 / (s_{.n} * t_{.p} / T) \end{aligned}$$

以下では煩雑を避けるために最初と最後の項だけ示します。

$$\begin{aligned} &= (f_{.11} - s_{.1} * t_{.1} / T)^2 / (s_{.1} * t_{.1} / T) + \dots \\ &+ (f_{.np} - s_{.n} * t_{.p} / T)^2 / (s_{.n} * t_{.p} / T) \\ &= [(f_{.11} * T - s_{.1} * t_{.1}) / T]^2 / (s_{.1} * t_{.1} / T) + \dots \\ &+ [(f_{.np} * T - s_{.n} * t_{.p}) / T]^2 / (s_{.n} * t_{.p} / T) \\ &= [(f_{.11} * T - s_{.1} * t_{.1})^2 / T^2] / (s_{.1} * t_{.1} / T) + \dots \\ &+ [(f_{.np} * T - s_{.n} * t_{.p})^2 / T^2] / (s_{.n} * t_{.p} / T) \end{aligned}$$

$$\begin{aligned}
&= [(f.11 * T - s.1 * t.1)^2 / T] / (s.1 * t.1) + \dots \\
&+ [(f.np * T - s.n * t.p)^2 / T] / (s.n * t.p) \\
&= [(f.11 * T - s.1 * t.1)^2] / (s.1 * t.1 * T) + \dots \\
&+ [(f.np * T - s.n * t.p)^2] / (s.n * t.p * T) \\
&= (f^2.11 * T^2 - 2 f.11 * T * s.1 * t.1 + s^2.1 * t^2.1) / (s.1 * t.1 * T) + \dots \\
&+ (f^2.np * T^2 - 2 f.np * T * s.n * t.p + s^2.n * t^2.p) / (s.n * t.p * T) \\
&= f^2.11 * T^2 / (s.1 * t.1 * T) - 2 f.11 * T * s.1 * t.1 / (s.1 * t.1 * T) + s^2.1 * t^2.1 / (s.1 * t.1 * T) \\
&+ \dots \\
&+ f^2.np * T^2 / (s.n * t.p * T) - 2 f.np * T * s.n * t.p / (s.n * t.p * T) + s^2.n * t^2.p / (s.n * t.p * T) \\
&= \boxed{f^2.11 * T / (s.1 * t.1)} - \boxed{2 f.11} + \boxed{s.1 * t.1 / T} \\
&+ \dots \\
&+ \boxed{f^2.np * T / (s.n * t.p)} - \boxed{2 f.np} + \boxed{s.n * t.p / T}
\end{aligned}$$

上の式の四角で囲った縦の列の3項を[1], [2], [3]とします。

$$\chi^2 = [1] - [2] + [3]$$

$$[1] = f^2.11 * T / (s.1 * t.1) + \dots + f^2.np * T / (s.n * t.p)$$

$$[2] = 2 f.11 + \dots + 2 f.np$$

$$[3] = s.1 * t.1 / T + \dots + s.n * t.p / T$$

$$[1] = f^2.11 * T / (s.1 * t.1) + \dots + f^2.np * T / (s.n * t.p)$$

$$= T [f^2.11 / (s.1 * t.1) + \dots + f^2.np / (s.n * t.p)]$$

$$= T \sum_{i:1,n} \sum_{j:1,p} f.i,j / (s.i * t.j)$$

$$[2] = 2 f.11 + \dots + 2 f.np$$

$$= 2 (f.11 + \dots + f.np)$$

$$= 2 T \leftarrow \text{上の表を参照}$$

$$[3] = s.1 * t.1 / T + \dots + s.n * t.p / T$$

$$= (s.1 * t.1 + \dots + s.n * t.p) / T$$

$i:1,n, j:1,p$ について見渡します。

$$= (s.1 * t.1 + s.1 * t.2 + \dots + s.1 * t.p$$

$$+ s.2 * t.1 + s.2 * t.2 + \dots + s.2 * t.p$$

...

$$+ s.n * t.1 + s.n * t.2 + \dots + s.n * t.p) / T$$

$$= [s.1 * (t.1 + t.2 + \dots + t.p)$$

$$+ s.2 * (t.1 + t.2 + \dots + t.p)$$

$$\begin{aligned}
& + \dots \\
& + s.n * (t.1 + t.2 + \dots + t.p) / T \\
& = [(s.1 + s.2 + \dots + s.n) * (t.1 + t.2 + \dots + t.p)] / T \\
& = T * T / T \quad \leftarrow \text{上の表を参照} \\
& = T
\end{aligned}$$

よって

$$\begin{aligned}
\chi^2 &= [1] - [2] + [3] \\
&= T \sum_{i:1,n} \sum_{j:1,p} f_{ij} / (s_i * t_j) - 2T + T \\
&= T \sum_{i:1,n} \sum_{j:1,p} f_{ij} / (s_i * t_j) - T \\
&= T [\sum_{i:1,n} \sum_{j:1,p} f_{ij} / (s_i * t_j) - 1]
\end{aligned}$$

この別式はカイ二乗値を扱った多くの統計学の本で扱われていませんが、複雑な期待値 e_{ij} を計算する必要がないので、この式を使えば Excel 上の作業（→後述）もプログラミングも簡単になります。

* 参考：伊藤雅光(2002)『計量言語学入門』大修館書店

● プログラム

次の関数プログラムは、カイ二乗値と確率を返します。別式(1)を使用しました。

```

Function Chi(a, b, c, d, sel) 'カイ二乗値
  Chi = (a + b + c + d) * (Abs(a * d - b * c)) ^ 2 _
    / ((a + b) * (c + d) * (a + c) * (b + d)) 'カイ二乗値 (sel=0)
  If sel = 1 Then Chi = Application.ChiSq_Dist_RT(Chi, 1) '確率 (sel=1)
End Function

```

9.9.3. 検定の評価

カイ二乗統計量は、期待値とのズレであるということを見ました。それではこの値がどの程度大きければ差があるといえるのでしょうか。2-2 の分割表では次のように決まっています。

有意水準 閾値	
5%	3.841
1%	6.634

閾値とは、カイ二乗統計量の値がそれ以下であれば成り立たないということです。あらためて先ほどの値は 4.288 なので、5%の閾値よりも大きいことになります。従って、この結果は「5%水準で有意な差がある」と解釈

できます。一方、1%水準の閾値は 6.634 なので、この水準では H_0 を棄却することはできません。

さきほど「2x2 の分割表では」という但し書きをつけましたが、この点は重要ですので触れておきます。カイ二乗統計量は期待値からのズレの合計であるということを見ましたが、マス目が増えれば増えるほど合計の値が大きくなります。たとえば、2x2 のマスと 4x4 のマスではマスの数は 4 マスと 16 マスですので、平等に扱うのはおかしいでしょう。つまり、有意水準の閾値の値も、マス目の数によって大きくなっていくということになります。

この基準は「自由度」(degree of freedom, df)とよばれます。自由度というのは自由に値を決めることができるマスの数のことです。たとえば、2x2 のマスでは、1 つのマスを決めると、縦と横の合計が同じならば他のすべてのマスの値は自動的に決まってしまうので自由度は 1 ということになります。次の表で方法 A の「効果がある」を 10 とすると、方法 A の「効果がない」は 84、方法 A の「効果がある」は 98、方法 B の「効果がない」は 4 に決まります。

実測値 1	効果がある	効果がない	
方法 A	10	94-10	94
方法 B	108-10	102-(108-10)	102
	108	88	196

なお、n 行 p 列のクロス集計表の自由度は、 $(n-1)(p-1)$ で求めることができます。以上のことをまとめて次のように表します¹⁷。

$$\chi^2 = 4.288 > \chi^2(\text{df: } 1, \text{ p: } 0.05) = 3.841$$

これは「カイ二乗統計量は 4.288 で、自由度 1 の場合の 5% 有意水準の 3.841 よりも大きく統計的に有意である」という意味です。

● Excel によるカイ二乗検定

Excel のシートで青色で示したデータ(Data)と有意水準(a)を書き込むとそれぞれの行和、列和、総和、カイ二乗値(χ^2)、閾値(Crit.value)、上側確率(Upper-prob.)、有意性判定(Significance)が出力されます。

¹⁷ df は自由度(degree of freedom)、p は確率(probability)を示します。

	A	B	C	D	E	F	G	H
1	Chi^2 test: 2 x 2							
2	Data:	1	2	Sum				
3	1	48	24	72	=SUM(B3:C3)			
4	2	32	36	68				
5	Sum	80	60	140				
6		=SUM(B3:B4)						
7	a	0.050						
8	Chi^2	5.490	=D5*(B3^2/D3/B5+C3^2/D3/C5+B4^2/D4/B5+C4^2/D4/C5-1)					
9	Critical value	3.841	=CHIINV(B7,1)					
10	Upper-prob.	0.019	=1-CHISQ.DIST(B8,1,1)					
11	Significance	S. (H1)	=IF(B10<B7,"S. (H1)", "NS (H0)")					
12								

カイ二乗値は上で説明した別式を使用しました。上側確率は確率の総和(1)からカイ二乗累積確率を引いて求めます。この上側確率がはじめに設定した有意水準以下であれば「有意」として対立仮説(H1)を採用します。

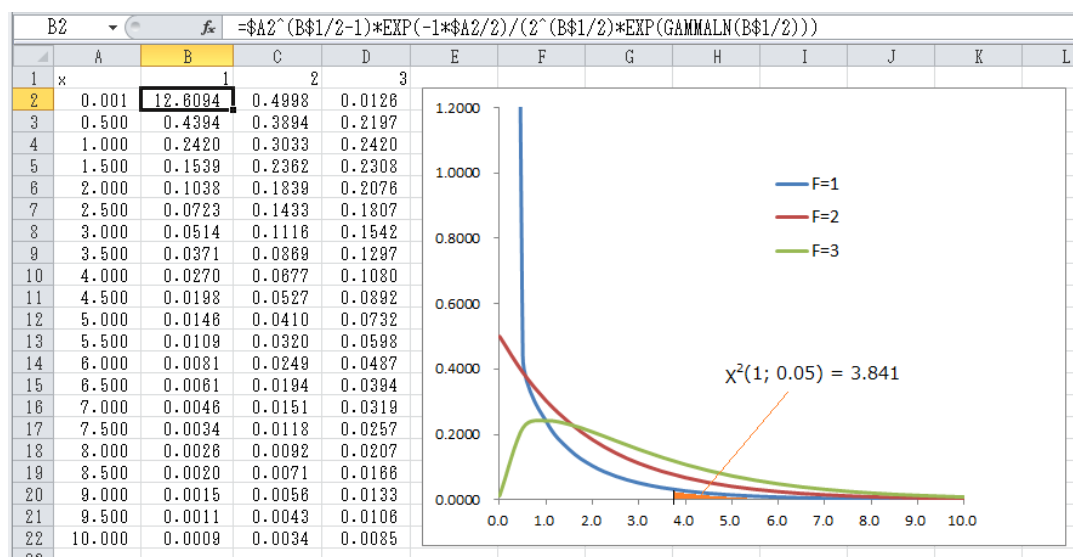
■カイ二乗分布の実験 (1)

カイ二乗の「限界値」は「有意水準」と「自由度」によって決まります。たとえば、自由度=1、有意水準=0.05ならば、限界値は3.841になります。

この限界値は非常に複雑な数式によって導かれるもので、これを理解することは困難です。次に示すシートは、カイ二乗分布を示す関数を自由度1, 2, 3について求めたものですが、セル[B2]の数式は、次のような関数を使います。このように非常に複雑な式なのです。

$$B2 = \$A2^{(B\$1/2-1)} * EXP(-1 * \$A2/2) / (2^{(B\$1/2)} * EXP(GAMMALN(B\$1/2)))$$

ここで自由度(F) = 1 の線の 3.841 (閾値) の値の右側の面積 (上側確率) が全体の 5%になることを示しています。



カイ二乗値の確率分布は実際に実験をすることによって納得することができます。はじめに次の不偏標本分散の分布に関する定理を確認します(倉田・星野 2009: 207)。

X_1, X_2, \dots, X_n を正規母集団 $N(m, v)$ からの無作為標本とすると、次の Y は自由度 $n - 1$ のカイ二乗分布に従う。

$$Y = 1 / v \sum_{i=1, n} (X_i - M)^2$$

(m, v は母集団の平均と分散, M は標本の平均と分散)

この定理が実際にどのように実現されるのかを見るために Excel で次の実験をしました。

	A	B	C	D	E	F	G	H	I
1	Chi ² test			Trial	a	b	c	d	chi ² .v
2	Set			1	7.0	10.4	9.2	10.0	1.331
3	mr	10.000		2	8.0	12.0	10.3	11.4	1.921
4	vr	5.000		3	6.7	10.8	13.1	6.4	6.300
5	n	4		4	11.3	6.0	8.9	9.5	2.983
6				5	8.8	12.0	11.3	11.5	1.249
7	Trial			6	9.4	10.3	13.5	10.3	1.901
8	m	9.999		7	11.8	11.0	11.1	9.7	0.441
9	v	5.041		8	7.6	8.4	11.0	6.8	1.947
10				9	10.4	14.6	8.3	9.8	4.285
11	Renew: [F9]			10	11.3	11.8	10.9	13.5	0.789
12				11	10.5	10.0	6.6	11.4	2.594

B3 と B4 に乱数の平均(mr)と分散(vr)を設定し、データ数は 4 とします。B8 と B9 は、範囲 E:H のすべての数値の平均(m)と分散(v)が表示されます。

$$B8=AVERAGE(E:H)$$

$$B9=VAR(E:H)$$

E2 のセルには次の式で範囲 [0, 1] の一様分布乱数 RAND() に対応する正規分布 (平均 m, 標準偏差 \sqrt{v}) に従う値が表示されます。

$$E2=NORMINV(RAND(), \$B\$3, SQRT(\$B\$4))$$

それを E 列から H 列までコピーし、その E:H の行を 2000 行コピーします。これで $2000 * 4 = 8000$ の数値が E:H の列に表示されます。

I2 のセルに $=DEVSQ(E2:H2)/\$B\9 を設定します。DEVSQ は次の式の偏差平方和

$$\sum_{(i=1, n)} (X_i - M)$$

を返します。これを B9 のセルにある母分散(v)で割った値になります。分母の母分散(v)は B9 の値です。

$$I2=DEVSQ(E2:H2)/\$B\$9$$

これを I 列全体にコピーします。

次に Q, R 列に 0 から 0.5 間隔の数値を用意し, S2 と T2 に次の式を入れます。

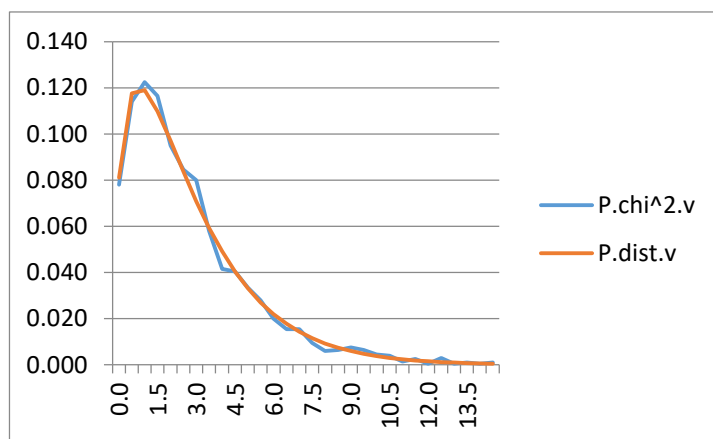
$$S2=COUNTIFS(\$I\$2:\$I\$2001,">"&Q2,\$I\$2:\$I\$2001,"<="&R2)/2000$$

$$T2=CHISQ.DIST(R2,3,1)-CHISQ.DIST(Q2,3,1)$$

S2 には条件 (Q2 より大きく R2 以下であれば数える) を満足する数値の数を全個数 2000 で割った値が表示されます。T2 には R2 に対応する自由度 3 の累積確率(1)から Q2 に対応する自由度 3 (n = 4) の累積確率(第 3 引数:1)を引いた数値が返されます。これが 0.0 から 0.5 の範囲にある確率の全体です。

Q	R	S	T
From	To	P.chi^2.v	P.dist.v
0.0	0.5	0.072	0.081
0.5	1.0	0.123	0.118
1.0	1.5	0.130	0.110

Excel の折れ線グラフを出力させると, 先の定理のとおり Y はカイ二乗分布に従っていることがわかります。



● 母分散の検定：外国語試験の点数のバラツキ

ここで, 次の分散(V)の式に注意します。

$$V = 1 / n \sum [i=1,n] (X.i - M)^2 \leftarrow \text{「統計量」 「分散」}$$

$$\rightarrow \sum [i=1,n] (X.i - M)^2 = n V$$

そこで先の Y の式は

$$\begin{aligned}
 Y &= 1 / v \sum [i=1,n] (X.i - M)^2 \\
 &= 1 / v n V \\
 &= n V / v
 \end{aligned}$$

よって、次の定理が確認できます(小寺 2002: 154-155)。

X.1, X.2, ..., X.n を正規母集団 N(m, v)からの無作為標本とすると、次の Y は自由度 n - 1 のカイ二乗分布に従う。

$$Y = n V / v$$

(v は母集団の分散, V は標本の分散)

この定理を使って「母分散の検定」が行われます。たとえば、過去の大量の外国語試験データ(これを母集団とします)の点数の分散(v)が 95 であり (sd = 9.747), 特定のテスト(標本数 n = 25)の点数の分散(V)が 50 であったとき(SD= 7.071), この分散(V)が有意に母分散(v)より小さいか(V < v), ということをも、有意水準 0.05 の下側確率で検定することにします。

$$\begin{aligned}
 y &= n V / v = 25 * 50 / 95 = 13.158 \\
 \text{CHISQ.DIST}(13.158, 25 - 1, 1) &= 0.037
 \end{aligned}$$

よって、帰無仮説(V = v)は棄却され、分散(V=95)は有意に大きい(対立仮説: V < v), と言えます(有意水準 0.05)。

	G	H		G	H
1	V-test (v:V). Result		1	V-test (v:V). Result	
2	v	95.000	2	v	95
3	sd	9.747	3	sd	=SQRT(H2)
4	N	25	4	N	25
5	V	50.000	5	V	50
6	SD	7.071	6	SD	=SQRT(H5)
7	Significance level	0.050	7	Significance level	0.05
8	Y-value	13.158	8	Y-value	=H4*H5/H2
9	Lower probability	0.037	9	Lower probability	=CHISQ.DIST(H8,H4-1,1)
10	Lower critical value	36.415	10	Lower critical value	=CHISQ.INV.RT(H7,H4-1)
11	Lower significance	S. (H1)	11	Lower significance	=IF(H9<H7,"S. (H1)", "NS (H0)")
12	Upper probability	0.963	12	Upper probability	=CHISQ.DIST.RT(H8,H4-1)
13	Upper critical value	13.848	13	Upper critical value	=CHISQ.INV(H7,H4-1)
14	Upper significance	NS (H0)	14	Upper significance	=IF(H12<H7,"S. (H1)", "NS (H0)")
15	Lower 2-tailed critical v.	39.364	15	Lower 2-tailed critical v.	=CHISQ.INV.RT(H7/2,H4-1)
16	Upper 2-tailed critical v.	12.401	16	Upper 2-tailed critical v.	=CHISQ.INV(H7/2,H4-1)
17	Two-tailed probability	0.073	17	Two-tailed probability	=MIN(H9,H12)*2
18	Two-tailed significance	NS (H0)	18	Two-tailed significance	=IF(H17<H7,"S. (H1)", "NS (H0)")

● プログラム

```
Function VarT(v, vp, n, sel)
'母分散の検定, sel=0(現確率); sel=1(下側); 2(上側); 3(両側)
  With Application
    If sel = 0 Then VarT = .ChiSq_Dist(n * v / vp, n - 1, 0) '現確率
    If sel = 1 Then VarT = .ChiSq_Dist(n * v / vp, n - 1, 1) '下側確率
    If sel = 2 Then VarT = .ChiSq_Dist_RT(n * v / vp, n - 1) '上側確率
    If sel = 3 Then '両側確率
      VarT = .Min(.ChiSq_Dist(n * v / vp, n - 1, 1), .ChiSq_Dist_RT(n * v /
vp, n - 1)) * 2
    End If
  End With
End Function
```

■ カイ二乗分布の実験 (2)

次に同様のことを 2*2 のクロス表のカイ二乗値について確かめましょう。

	J	K	L	M	N	O
v	a+b	c+d	a+c	b+d	T	chi ² .e
2	16.7	23.5	19.8	20.4	40.2	0.005
7	20.4	22.3	18.7	24.0	42.8	0.122
6	22.8	20.8	18.0	25.6	43.6	0.267
8	40.7	33.8	30.2	33.4	43.6	4.200

$$J2=E2+F2$$

$$K2=G2+H2$$

$$L2=E2+G2$$

$$M2=F2+H2$$

$$N2=\text{SUM}(E2:H2)$$

$$O2=N2*(E2^2/J2/L2+F2^2/J2/M2+G2^2/K2/L2+H2^2/K2/M2-1)$$

O2 にはカイ二乗値の別式を使用しました (→前述「カイ二乗値の簡単な別式」)。

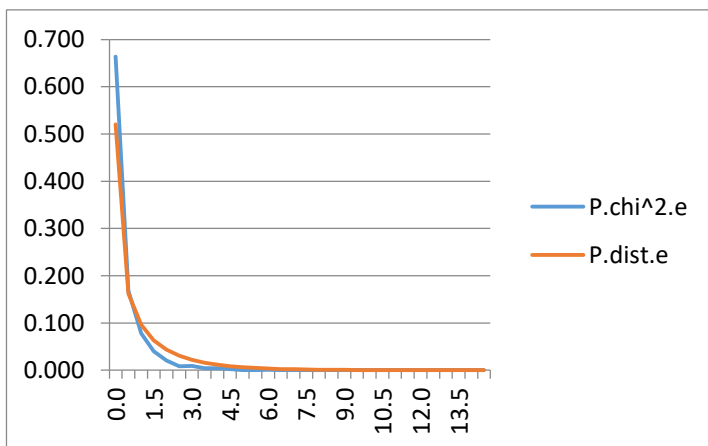
P	Q	R	S	T	U	V	W
	From	To	P.chi ² .v	P.dist.v	P.chi ² .e	P.dist.e	
	0.0	0.5	0.071	0.081	0.656	0.520	
	0.5	1.0	0.126	0.118	0.172	0.162	
	1.0	1.5	0.119	0.119	0.078	0.097	
	1.5	2.0	0.118	0.110	0.040	0.063	

先と同様に U2 と V2 にカイ二乗の実験で得た実測値(P.chi².e)と Excel 関数で求めた理論値(P.dist.e)を求めます。V2 の自由度は 1 です。

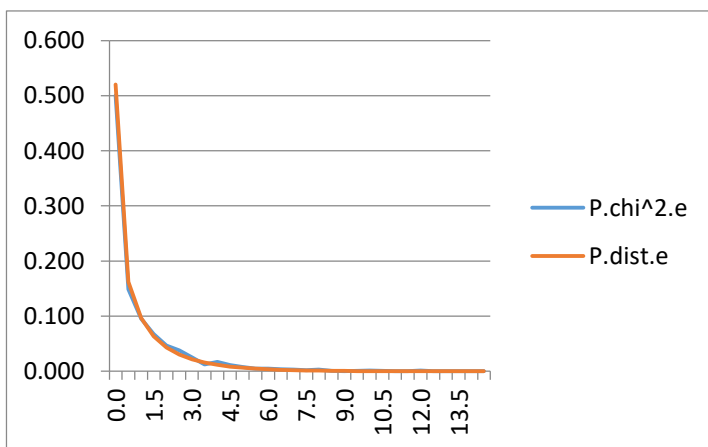
$$U2 = \text{COUNTIFS}(\$O\$2:\$O\$2001, ">"\&Q2, \$O\$2:\$O\$2001, "<=" \&R2) / 2000$$

$$V2 = \text{CHISQ.DIST}(R2, 1, 1) - \text{CHISQ.DIST}(Q2, 1, 1)$$

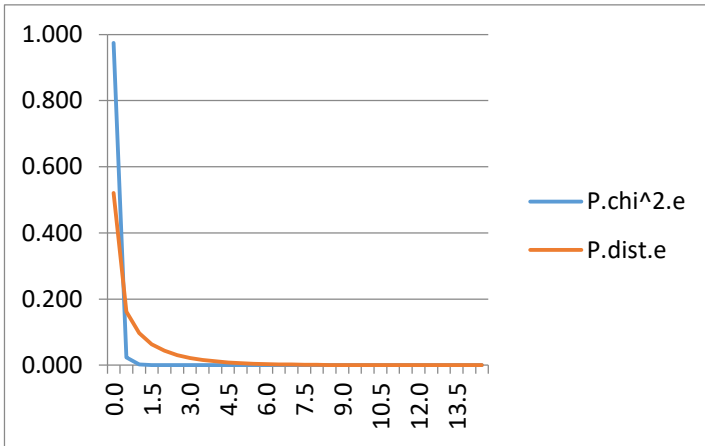
次の図はカイ二乗の実験で得た実測値(P.chi².e)と Excel 関数で求めた理論値(P.dist.e)の確率分布を示します。



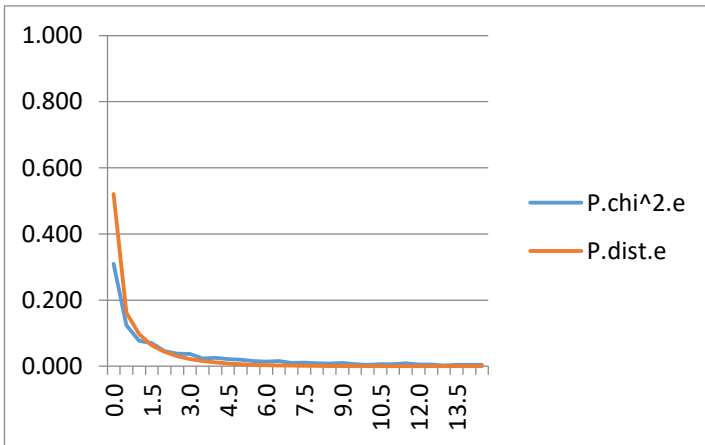
このように実測値と理論値は近似していますが若干の差異がみられます。この図は正規分布乱数の平均を 10, 分散を 5 とした場合の確率分布です。分散と平均を同値(たとえば 10)にすると, 実測値と理論値はほとんど完全に一致します。



一方, 分散を極端に少なくし, $m=10, v=1$ の正規分布乱数で実験すると実測値と理論値は大きく乖離します。平均からのブレがほとんど生じていないデータです(標準偏差 $SD=1$)。



逆に分散を極端に大きくし， $m=10$ ， $v=25$ の正規分布乱数で実験すると，上下関係が逆転して，やはり乖離します。これは平均から大きくブレが生じるようなデータです（標準偏差 $SD=5$ ）。



一般にカイ二乗値は標本の母集団が正規分布（平均 m ，分散 v ）に従うことを前提にしています。クロス集計表のカイ二乗値は，ポアソン分布になるデータのように平均(m)=分散(v)のとき実測値と理論値がほとんど完全に一致することを見ました。一方，平均と分散が大きく乖離するような正規分布になる集団を，標本の母集団として前提にすることができません。カイ二乗検定では「データが平均値と分散値に近い正規分布に従う母集団からの標本とすれば」という前提をおく必要があるでしょう。

9.10. 正確検定

カイ二乗検定は，次のそれぞれのセルが頻度 5 以上でないと正確な検定ができないことが知られています。

X[準備]:Y[判断]	Y(m)	Y(t)	計
X(m)	$a(+:+) = 3$	$b(+:-) = 1$	$e = a+b = 4$

X(t)	c(-:+) = 1	d(-:-) = 3	f = c+d = 4
計	g = a+c = 4	h = b+d = 4	n = a+b+c+d = 8

そして、カイ二乗検定の判断に使われるカイ二乗累積分布を導く式を理解するには高度な数学的準備が必要です。

ここでは、わかりやすい Fisher の「正確検定」(Fisher's exact test)を取り上げます。必要な数学的準備は高校数学の範囲内の「組み合わせ」(combination)だけです。

Agresti (1996, 2003: 54-56)は Fisher の「紅茶の実験」の例を挙げて正確検定を説明しています。それによると、Fisher の同僚は紅茶を入れたところを見なくても試飲することによってミルクと紅茶のどちらを先に入れたのかがわかる、と言ったそうです。そこで、次のような実験をしました。8個のカップを用意し、そのうち、4個のカップは先にミルクを入れ、4個のカップには先に紅茶を入れて、同僚にそのことを伝えてありました。

上の表の X(m)はミルク(m)を先に入れたカップの数を示し、X(t)は紅茶を先に入れたカップの数を示します。Y(m)は同僚がミルクが先だと判断したカップの数であり、Y(t)は紅茶が先だと判断したカップの数を示します。a(+:+) = 3 であり、d(-:-) = 3 という数は同僚がかなりの精度でミルクと紅茶の順番を当てているようです¹⁸。しかし、上の数値は偶然でも起こることかも知れません。そこで、次の帰無仮説と対立仮説を設定します。

帰無仮説：同僚の順番の判断(Y)が実際の順番(X)と独立している

対立仮説：同僚の順番の判断(Y)と実際の順番(X)は連関している

はじめに 8 個のカップ(c1, c2, ..., c8)からミルクを先に入れた X(m)のカップ 4 個を選ぶ組み合わせ(combination: C)を数えます。n 個のカップから r 個のカップを取り出す場合の数は、順列(permutation: P)と組み合わせ(combination: C)を使うと

$${}_n C_r = {}_n P_r / r! = [n! / (n - r)!] / r! = n! / [r! (n - r)!]$$

具体的な数値をあてはめると

$$\begin{aligned} {}_8 C_4 &= 8! / (4! * (8-4)!) = 8! / (4! * 4!) \\ &= (8 * 7 * 6 * 5 * 4 * 3 * 2 * 1) / [(4 * 3 * 2 * 1) * (4 * 3 * 2 * 1)] \\ &= (8 * 7 * 6 * 5) / (4 * 3 * 2 * 1) = 7 * 2 * 5 = 70 \end{aligned}$$

この 70 が全体の組み合わせの場合の数を示します。

次に、Y(m)の計 4 個のカップ {c1, c2, c3, c4} から正しく当てた 3 個のカップを取り出す場合の数は

¹⁸ 実際に同僚はすべてを正しく判定したそうなので、[a, b, c, d]は[4, 0, 0, 4]になったはずですが、ここでは仮に[3, 1, 1, 3]として説明します。

$${}_4C_3 = 4! / (3! * 1!) = 4 \text{ 通り}$$

そして Y(t)の計 4 個のカップ {c5, c6, c7, c8}から順番を間違えた 1 個 (b)のカップを取り出す場合の数は

$${}_4C_1 = 4! / (1! * 3!) = 4 \text{ 通り}$$

4 個のカップ {c1, c2, c3, c4}から正しく当てた 3 個 (a)のカップを取り出し、かつ、4 個のカップ {c5, c6, c7, c8}から順番を間違えた 1 個 (b)のカップを取り出す取り出し方の数は

$${}_8C_4 * {}_4C_1 = 4 \text{ 通り} * 4 \text{ 通り} = 16 \text{ 通り}$$

一方、全体のカップの数 8 個から 4 個を取り出す取り出し方の数は、先に見たように、全部で

$${}_8C_4 = 70 \text{ 通り}$$

よって上の表のような数字 [a, b, c, d]の分布の生起確率(P)は

$$P = {}_4C_3 * {}_4C_1 / {}_8C_4 = 4 * 4 / 70 = 16 / 70 = .229$$

この確率は上の表の場合 [3, 1, 1, 3]だけを取り出した確率ですが、ここで可能なすべてのケースの確率を求めます。次の表は a, b, c, d のそれぞれの数値の可能性と、その生起確率(P), 累積確率(CP)を示します。

a	b	c	d	P	CP
0	4	4	0	${}_4C_0 * {}_4C_4 / {}_8C_4 = 1 * 1 / 70 = .014$	1.000
1	3	3	1	${}_4C_1 * {}_4C_3 / {}_8C_4 = 4 * 4 / 70 = .229$.986
2	2	2	2	${}_4C_2 * {}_4C_2 / {}_8C_4 = 6 * 6 / 70 = .514$.757
3	1	1	3	${}_4C_3 * {}_4C_1 / {}_8C_4 = 4 * 4 / 70 = .229$.243
4	0	0	4	${}_4C_4 * {}_4C_0 / {}_8C_4 = 1 * 1 / 70 = .014$.014

上の表の累積確率(CP)は、それぞれの確率以上の確率を足し合わせた数値です。ここで確率(P)の総和(.014 + .229 + ... + .014)は 1.000 になることを確認しましょう。

さて、問題の、[a, b, c, d] = [3, 1, 1, 3]のケースの生起確率(.229)と、さらに珍しいケース [4, 0, 0, 4]の生起確率(.014)を足し合わせると .243 になることから、帰無仮説を棄却する強い根拠にはなりません。5%以下の確率にならないければ帰無仮説を棄却することはできないからです。検定に使われる累積確率は小さいほど対立仮説を支持します。

* Fisher の正確検定については次を参照しました。

Agresti, A. 1996. An Introduction to Categorical Data Analysis. 渡邊・菅波・

吉田・角野・寒水・松永訳 2003 『カテゴリカルデータ解析入門』 サイエンティスト社

S. Aoki 2016.12.4. <http://aoki2.si.gunma-u.ac.jp/lecture/Cross/Fisher.html>

B.S. Everitt. 1977. *The Analysis of Contingency Table*, Chapman & Hall/CRC, pp.15-18. (片側検定)

市原清志 1990 『バイオサイエンスの統計学』 南江堂

水本篤 2010 「サンプルサイズが小さい場合の統計的検定の比較ーコーパス言語学・外国語教育学への適用」 統計数理研究所共同研究レポート 238 『言語コーパス分析における統計的処理手法の検討』 pp.1.14

H. Okumura 2016.12.3.

<https://oku.edu.mie-u.ac.jp/~okumura/stat/fishertest.html>

芝祐順・南風原朝和 1990 『行動科学における統計解析法』 東京大学出版会, pp.204-206.

■ 限定された資料の分析

次の表は, 言語現象(L: たとえば漢字の使用頻度: 使用 L(+), 不使用 L(-)) のテキスト T.1, T.2 中の分布を示します。

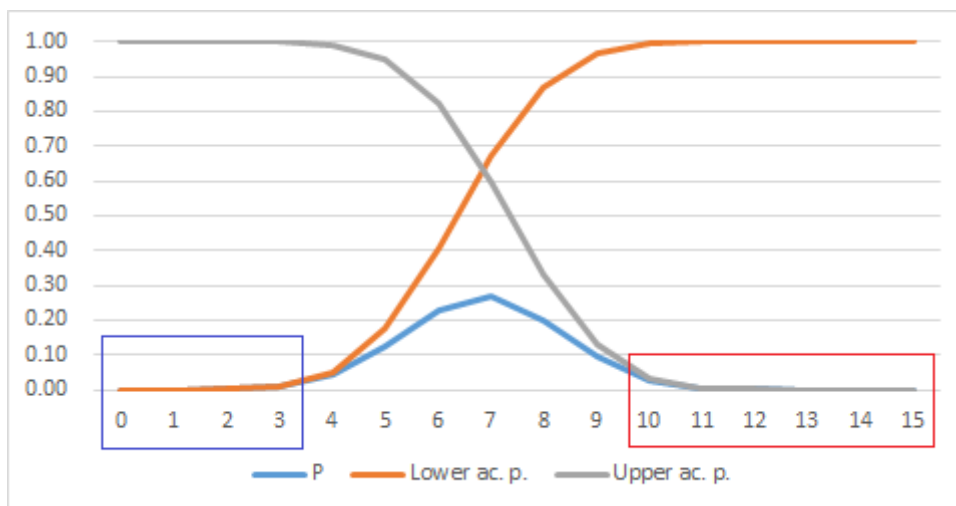
L.T.	T.1	T.2	計	当該確率(t)	0.0287
L(+)	a: 10	b: 5	e:15	下側累積確率(p)	0.9942
L(-)	c: 6	d: 14	f:20	上側累積確率(q)	0.0345
計	g:16	h:19	N:35	両側累積確率(r)	0.0442

この分析で L(+), L(-) と T.1, T.2 だけを取り上げて, 漢字の頻度 L(+) が T.1 と T.2 の間で有意に異なるかどうかを検証します。

a	b	c	d	N	Fisher. p.	Lower p.	Upper p.
0	15	16	4	35	0.0000	0.0000	1.0000
1	14	15	5	35	0.0001	0.0001	1.0000
2	13	14	6	35	0.0010	0.0011	0.9999
3	12	13	7	35	0.0087	0.0097	0.9989
4	11	12	8	35	0.0424	0.0521	0.9903
5	10	11	9	35	0.1242	0.1763	0.9479
6	9	10	10	35	0.2278	0.4041	0.8237
7	8	9	11	35	0.2662	0.6703	0.5959
8	7	8	12	35	0.1997	0.8700	0.3297
9	6	7	13	35	0.0956	0.9655	0.1300
10	5	6	14	35	0.0287	0.9942	0.0345
11	4	5	15	35	0.0052	0.9994	0.0058

12	3	4	16	35	0.0005	1.0000	0.0006
13	2	3	17	35	0.0000	1.0000	0.0000
14	1	2	18	35	0.0000	1.0000	0.0000
15	0	1	19	35	0.0000	1.0000	0.0000

ここで「当該確率(t)」は順列・組み合わせを使って計算した確率を示し、「下側累積確率(p)」は、当該の確率までの累積です。この累積確率を、下側の片側検定（有意に小さい数値であるかを検証）に使用します(次のグラフの赤線で囲まれた部分)。次に「上側累積確率(q)」は、当該確率以上のすべての確率の累積です。この確率は、上側の片側検定（有意に大きい数値であるかを検証）に使用します(グラフの赤線で囲まれた部分)。最後の「両側累積確率(r)」は当該確率よりも小さな確率（下側と上側）の累積です(上表の P(t)の太字部分)。この確率は、上側の片側検定（有意に異なる数値であるかを検証）に使用します。Fisher の正確検定では、ふつう下側確率と上側確率の幅が対照的に同じになったり、両側累積確率が片側累積確率の 2 倍にはなりません¹⁹。



● 超幾何確率

一般に次のような a, b, c, d の数値から次の式を用いて計算された確率は「超幾何確率」(hypergeometric probability: H)と呼ばれています。

X:Y	Y+	Y-	sum
X+	a (+: +)	b (+: -)	e = a+b
X-	c (-: +)	d (-: -)	f = c+d
sum	g = a+c	h = b+d	N = a+b+c+d

¹⁹ 下側確率と上側確率の幅が対照的に同じになり、両側累積確率が片側累積確率の 2 倍になるのは、Fisher の「紅茶の実験」に示されているように周辺度数がすべて場合(したがって a = d, b = c)です。

$$H(a, b, c, d) = {}_{a+c}C_a * {}_{b+d}C_b / {}_{a+b+c+d}C_{a+b}$$

この H を， 次の変数を使って整理しましょう。

$$e = a+b, f = c+d, g = a+c, h = b+d, N = a+b+c+d$$

$$\begin{aligned} H(a, b, c, d) &= {}_gC_a * {}_hC_b / {}_nC_e \\ &= [g! / (g-a)! a!] * [h! / (h-b)! b!] / [N! / (N-e)! e!] \\ &= (g! / a! c!) * (h! / b! d!) / (N! / e! f!) \quad \leftarrow \text{上の表を参照} \\ &= (g! h! / a! b! c! d!) / (N! / e! f!) \\ &= (e! f! g! h!) / (a! b! c! d! N!) \end{aligned}$$

● プログラム

はじめに階乗の対数の配列 Ar を準備します(→二項検定のプログラム)。Fisher の正確検定の式の中で， e (=a+b), f (=c+d) , g (=a+c), h (=b+d), n (=a+b+c+d)は固定されているので， Log(e!)+ Log(f!)+ Log(g!)+ Log(h!)-Log(n!)を定数(k)とします。

```
Function FisherT(a, b, c, d, Optional sel)
'VBA: Fisher's Exact Test: sel=0(現確率); sel=1(下側); 2(上側); 3(両側)
  Dim Ar, k, s, f1, f2, i, t: ReDim Ar(a + b + c + d): Ar(0) = 0
  For i = 1 To a + b + c + d
    Ar(i) = Ar(i - 1) + Log(i) '階乗の対数の配列
  Next i
  k = Ar(a + b) + Ar(c + d) + Ar(a + c) + Ar(b + d) - Ar(a + b + c + d) '定数
  f1 = IIf(a > d, a - d, 0): f2 = IIf(a + b < a + c, a + b, a + c) '開始: 終了
  If sel = 0 Then f1 = a: f2 = a
  If sel = 1 Then f2 = a
  If sel = 2 Then f1 = a
  s = Exp(k - Ar(a) - Ar(b) - Ar(c) - Ar(d)) '現確率
  For i = f1 To f2
    t = Exp(k - Ar(i) - Ar(a + b - i) - Ar(a + c - i) - Ar(d - a + i)) '当該確率
    If sel < 3 Then FisherT = FisherT + t '累積確率
    If sel = 3 And Int(t * 10 ^ 7) <= Int(s * 10 ^ 7) Then FisherT = FisherT + t
      '両側累積確率
  Next
End Function
```

次は Excel 関数 HypGeomDist(a, a+c, a+b, a+b+c+d)を用いたプログラムです。HypGeomDist 関数がブラックボックスになりますが，出力は上のプログラムと同じです。

```

Function FisherU(a, b, c, d, sel) 'Excel.HypGeomDist(a, a+c, a+b, a+b+c+d)
'VBA: Fisher's Exact Test: sel=0(現確率); sel=1(下側); 2(上側); 3(両側)
  Dim s, f1, f2, i, t
  f1 = IIf(a > d, a - d, 0): f2 = IIf(a + b < a + c, a + b, a + c) '開始: 終了
  If sel = 0 Then f1 = a: f2 = a
  If sel = 1 Then f2 = a
  If sel = 2 Then f1 = a
  s = Application.HypGeomDist(a, a + c, a + b, a + b + c + d) '現確率
  For i = f1 To f2
    t = Application.HypGeomDist(i, a + c, a + b, a + b + c + d) '当該確率
    If sel < 3 Then FisherU = FisherU + t '累積確率
    If sel = 3 And Int(t * 10 ^ 7) <= Int(s * 10 ^ 7) Then FisherU = FisherU + t
    '両側累積確率
  Next
End Function

```

● カイ二乗検定における Yates の補正と Fisher の正確検定

小寺(2002: 181-182)は次の例によって、2*2表のカイ二乗値を補正する考え方 (Yates の補正) を説明しています。

A.B	B1: 罹病	B2: 非罹病	計
A1: 注射した	3	285	288
A2: 注射しない	7	175	182
計	10	460	470

このように小さな数値がある場合、わずかに1だけの増減がカイ二乗検定の結果を変えることがあります。そこで慎重を期すために、次の Yates の補正が適用されます(同: 181)。

周辺度数をこのままにして、各度数 a, b, c, d を四捨五入によって得られた整数と考え、本来 A, B が独立なのに H₀ [帰無仮説] が棄却されていもう可能性を少なくするように、a, b, c, d を 0.5 ずつ補正しようというのが、イエツの補正の考え方です。

具体的には、カイ二乗値の a に対応する値を

$$\text{chi}(a) = (a - a')^2 / a'$$

の代わりに

$$\text{chi.y}(a) = (|a - a'| - 0.5)^2 / a'$$

とします。これを a, b, c, d についてすべて足し上げた数値が Yates の補正をしたカイ二乗値になります。その全体は

$$\chi^2 = \frac{(a - a')^2}{a'} + \frac{(b - b')^2}{b'} + \frac{(c - c')^2}{c'} + \frac{(d - d')^2}{d'} \quad \dots(1)$$

$$\chi^2 = N (ad - bc)^2 / (e f g h) \quad \dots(2)$$

←参照：カイ二乗値の別式(1)

は、次のようになります。

$$\chi^2_{\text{yatee}} = N (|ad - bc| - 0.5 N)^2 / (e f g h) \quad \dots(2)$$

たしかにカイ二乗分布曲線は連続しているのに、度数は整数値なので、それを四捨五入によって得られた整数と考え、度数から 0.5 を引くことによって厳しく検定することは合理的に思えます。一般に 2*2 表であればすべて Yates の補正を適用するという考え方もあるようです。一方、統計学書の中には一切 Yates の補正に言及していない本もあります。

市原(1990: 138-139)によれば「最近の欧米の統計書ではその見解を否定しているものが多い」と述べ、「行かつ列の周辺度数が固定していると考えてよい場合には Fisher の方法で両側確率を計算するか Yates の補正による χ^2 値を用い、いずれも固定していないと考える場合には(通常はこちら)、無補正の χ^2 値を用いる(または経験的に Fisher の方法で片側確率を用いる)のが確率論的に妥当と考えられる」と述べています(同: 139)。

ここで重要なことは、周辺度数(上表の計)が固定されているかいないか、という判断です。小寺(引用箇所)の冒頭にも「周辺度数をこのままにして」と明記されているように、Yates の補正をしたカイ二乗検定と Fisher の正確検定は、周辺度数が固定されていることが条件になります。たしかに Fisher の「紅茶の実験」でも周辺度数は固定されていました。私たちの言語研究でも、限定されたテキストを母集団として対象にするときは周辺度数が固定されています。一方、一般のアンケート調査では母集団からのサンプリングになるので周辺度数は自由である、とも考えられますが、そのアンケート調査だけに限定すれば、周辺度数は固定されていることになります。

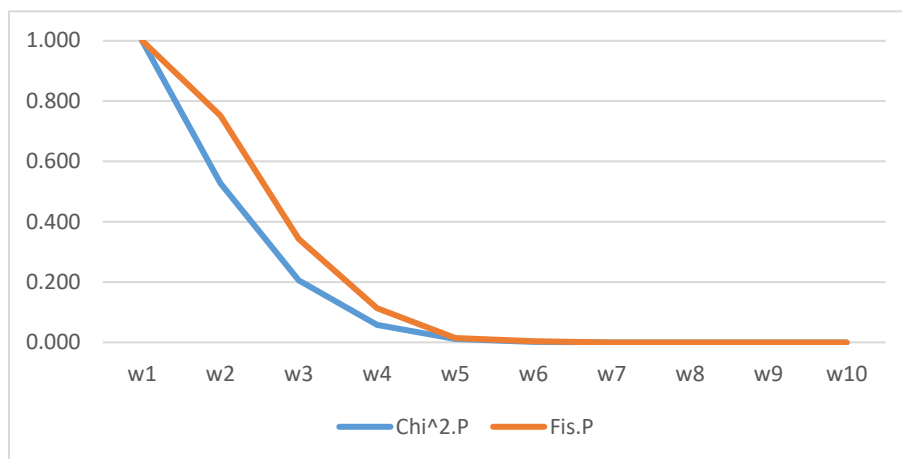
また、以下で見るように Yates の補正自体に疑義があるので、 a, b, c, d の小さな数値があつて、周辺度数が固定されているときは(限定されたインフォーマントや限定されたテキストを対象とする研究など)、Yates の補正をしたカイ二乗検定ではなく、Fisher の正確検定を適用すべきです。

● Yates の補正の問題と Fisher の正確検定

はじめに Yates の補正が適用されるはずの周辺度数を固定した例を見ま

す。次の表の a を順次増加させ、N を 40 に固定することにより周辺度数はすべて同じになります(=20)。そのときのカイ二乗値(χ^2)は、当然順次上昇します。それに伴って、カイ二乗値に対応する上側確率が下がります。それと Fisher の正確検定の両側確率を比較します。グラフを参照してください。

Word	a	b	c	d	N	Chi ² .a	Chi ² .b	Chi ² .c	Chi ² .d	chi ²	Chi ² .P	Fis.P
w1	10	10	10	10	40	0.000	0.000	0.000	0.000	0.000	1.000	1.000
w2	11	9	9	11	40	0.100	0.100	0.100	0.100	0.400	0.527	0.752
w3	12	8	8	12	40	0.400	0.400	0.400	0.400	1.600	0.206	0.343
w4	13	7	7	13	40	0.900	0.900	0.900	0.900	3.600	0.058	0.113
w5	14	6	6	14	40	1.600	1.600	1.600	1.600	6.400	0.011	0.015
w6	15	5	5	15	40	2.500	2.500	2.500	2.500	10.000	0.002	0.004
w7	16	4	4	16	40	3.600	3.600	3.600	3.600	14.400	0.000	0.000
w8	17	3	3	17	40	4.900	4.900	4.900	4.900	19.600	0.000	0.000
w9	18	2	2	18	40	6.400	6.400	6.400	6.400	25.600	0.000	0.000
w10	19	1	1	19	40	8.100	8.100	8.100	8.100	32.400	0.000	0.000

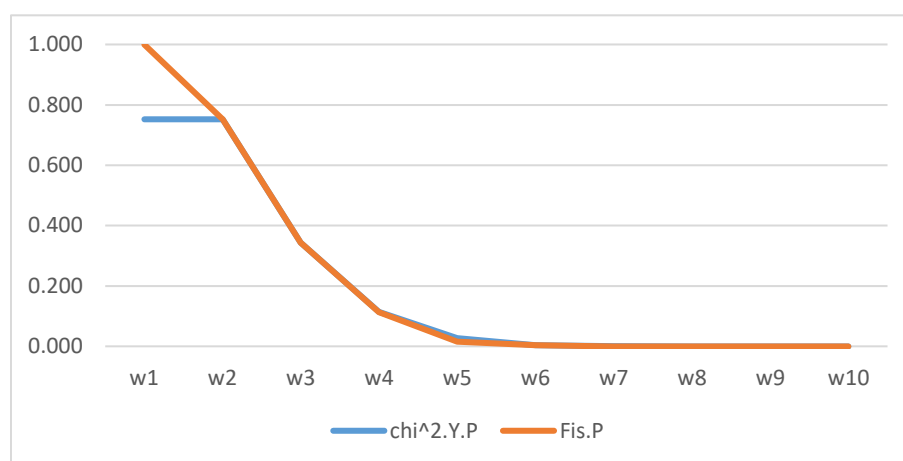


たしかに、カイ二乗値に対応する上側確率と Fisher の正確検定の両側確率にはかなりの乖離があります。そこで、周辺度数を固定したときにはカイ二乗検定ではなく、Fisher の正確検定を使うべきでしょう²⁰。

次に Yates の補正をしたカイ二乗値を使って同じ実験をします。

²⁰ さらに、Fisher の正確検定には片側検定も可能である、という利点があります。数値の多寡を評価するのなら、むしろ片側検定のほうがよいはずです。

Word	a	b	c	d	N	+Y.a	+Y.b	+Y.c	+Y.d	chi^2.Y	chi^2.Y.P	Fis.P
w1	10	10	10	10	40	0.025	0.025	0.025	0.025	0.100	0.752	1.000
w2	11	9	9	11	40	0.025	0.025	0.025	0.025	0.100	0.752	0.752
w3	12	8	8	12	40	0.225	0.225	0.225	0.225	0.900	0.343	0.343
w4	13	7	7	13	40	0.625	0.625	0.625	0.625	2.500	0.114	0.113
w5	14	6	6	14	40	1.225	1.225	1.225	1.225	4.900	0.027	0.015
w6	15	5	5	15	40	2.025	2.025	2.025	2.025	8.100	0.004	0.004
w7	16	4	4	16	40	3.025	3.025	3.025	3.025	12.100	0.001	0.000
w8	17	3	3	17	40	4.225	4.225	4.225	4.225	16.900	0.000	0.000
w9	18	2	2	18	40	5.625	5.625	5.625	5.625	22.500	0.000	0.000
w10	19	1	1	19	40	7.225	7.225	7.225	7.225	28.900	0.000	0.000



上の表と図を見ると、Yates の補正をしたカイ二乗値に対応する確率は Fisher の両側検定の確率に強く近似していることがわかります。しかし、ここで注意したいのは、最初の行(w1)です。ここでは、補正カイ二乗値が Fisher の両側検定の確率から大きく乖離して、さらに、w1 と w2 の間に変化が見られません。その理由は Yates の補正式を見ればわかります。

$$\text{chi.y(a)} = (|a - a'| - 0.5)^2 / a'$$

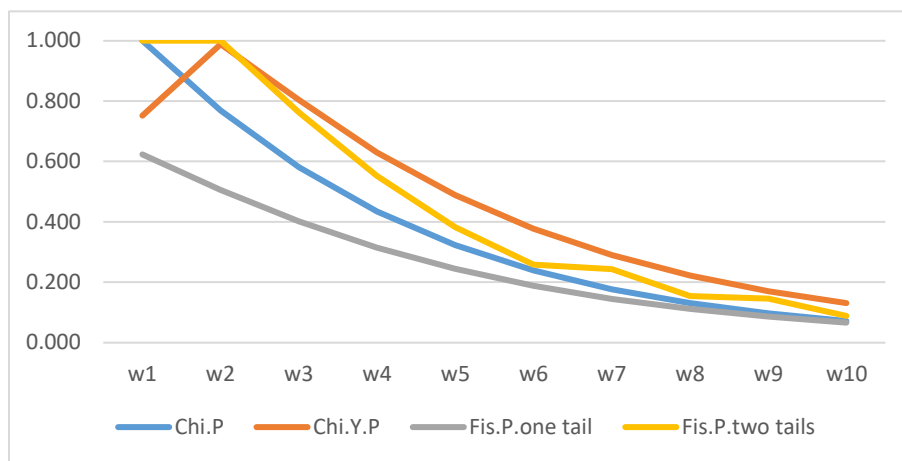
ここで a は実測値、a' は期待値を指します。上の式を見ると、a の補正カイ二乗値は a と a' の差ではなく、それからさらに 0.5 を引いて、期待値からの逸脱の程度を減らしていることがわかります。逸脱の程度が小さくなるので、その結果、確率が高くなり、検定が厳しくなるのです。

さて、最初の行では実測値と期待値が同じなので(=10)、上式の絶対値の中はゼロであり、分子は $0.5^2 = 0.25$ になります。分母は期待値(=10)なので、補正カイ二乗値は 0.025 になります。ところが、第 2 行(w2)を見ると、a = 11 であり、期待値は 10 なので、絶対値の中は 1 になります。このとき、分子

は $(1 - 0.5)^2 = 0.25$ になって、先の例(w1)と同じ結果になってしまうのです²¹。この原因は a がゼロ (0) でも 1 でも分子は同じになるからです。このことは、実測値が期待値と一致するときだけでなく、0.5 以下の違いであれば、やはり同様な問題を起こします。そしてデータはふつう平均値（期待値）の周りに集まりやすいので、このようなケースの発生率は高いと思われます。

次に、周辺度数が変化するときのカイ二乗検定，Yates 補正付きカイ二乗検定，Fisher 片側（上側）検定，Fisher 両側検定の確率(p 値)を比較しましょう。

Word	a	b	c	d	N	Chi.P	Chi.Y.P	Fis.P.one-tail	Fis.P.two-tails
w1	10	10	10	10	40	1.000	0.752	0.624	1.000
w2	12	10	10	10	42	0.768	0.988	0.506	1.000
w3	14	10	10	10	44	0.580	0.804	0.402	0.762
w4	16	10	10	10	46	0.434	0.629	0.315	0.552
w5	18	10	10	10	48	0.322	0.488	0.244	0.382
w6	20	10	10	10	50	0.239	0.377	0.188	0.258
w7	22	10	10	10	52	0.176	0.290	0.145	0.243
w8	24	10	10	10	54	0.130	0.222	0.111	0.154
w9	26	10	10	10	56	0.096	0.170	0.086	0.146
w10	28	10	10	10	58	0.071	0.130	0.066	0.088



ここでは、Yates 補正つきカイ二乗検定の p 値(Chi.Y.P)が w2 で一度上昇しています。この理由は w2 で実測値と期待値が a, b, c, d ですべて差が 0.5 になるため、補正(-0.5)によって差がゼロになっているためです。w1 と w3 ではそのようなことがないので、どちらでもわずかな差が生まれ、高い確率を生んでいます。しかし、w1, w3 よりも w2 のほうが確率が高くなるのは合理的ではありません。

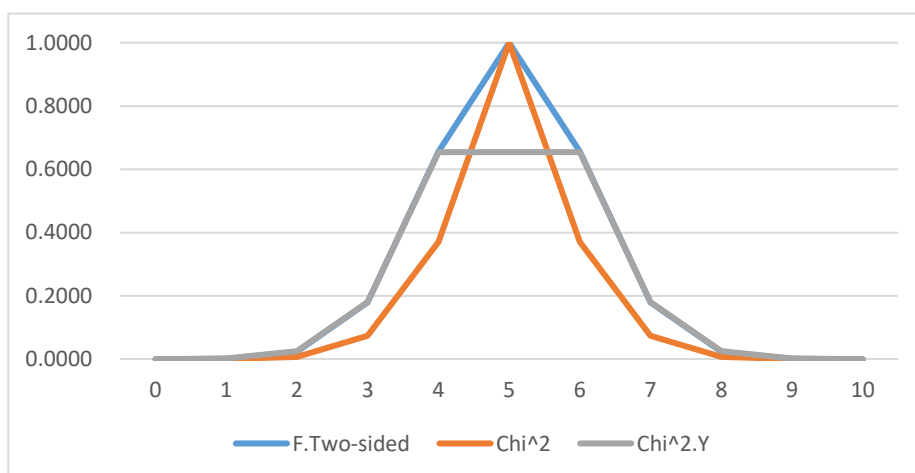
上の図を見ると、Yates の補正をしないカイ二乗の p 値と Fisher の片側

²¹ $(0 - 0.5)^2 = (1 - 0.5)^2$

検定の p 値は単調減少をしているので安定しています。市原(1990: 139)は、周辺度数の「いずれも固定していないと考える場合には(通常はこちら)、無補正の χ^2 値を用いる(または経験的に Fisher の方法で片側確率を用いる)のが確率論的に妥当と考えられる」と述べています。私はこの意見に基本的に賛成します。さらに、無補正の χ^2 値と Fisher の方法の片側確率を比較すると、Fisher の方法のほうが有用だと思います。なぜならば、(修正の有無にかかわらず) χ^2 値には片側(上側と下側)の観点がなく、データ全体(a, b, c, d)の分布の特異性(普通ではないこと)を調べることだけしかできませんが、Fisher の片側検定ならば、データの数値が普通(偶然)よりも特異に大きいのか、特異に小さいのか、ということも見極められるからです。また、これまでの実験と次の実験からわかるように、Fisher の検定(片側・両側)を使えば 2 x 2 の昇目の a に注目して、その有意性を検定することができます。

Fisher の正確両側検定 F.Two-sided とカイ二乗検定(Yates 補正なし Chi², あり Chi².Y) を次のデータで比較します。

a	b	c	d	N	F.Two-sided	Chi ²	Chi ² .Y
0	10	10	0	20	0.0000	0.0000	0.0001
1	9	9	1	20	0.0011	0.0003	0.0017
2	8	8	2	20	0.0230	0.0073	0.0253
3	7	7	3	20	0.1789	0.0736	0.1797
4	6	6	4	20	0.6563	0.3711	0.6547
5	5	5	5	20	1.0000	1.0000	0.6547
6	4	4	6	20	0.6563	0.3711	0.6547
7	3	3	7	20	0.1789	0.0736	0.1797
8	2	2	8	20	0.0230	0.0073	0.0253
9	1	1	9	20	0.0011	0.0003	0.0017
10	0	0	10	20	0.0000	0.0000	0.0001



上の表と図を見ると, Fisher の正確両側検定(F.Two-sided)は Yates 補正つきカイ二乗検定($\chi^2.Y$)と近似していること (グラフでは大部分が重なっている), そしてカイ二乗検定(Yates 補正あり $\chi^2.Y$)が $a = 5$ で補正の問題を起こして頂点が欠損していることがわかります。

● $n * p$ の表の正確検定

先に見たように, Fisher の正確検定は $a(+:+)$ の値に注目し, その有意性を検定します。もちろん $b(+:-)$, $c(-: +)$, $d(-:-)$ の値に注目することもできますが, いずれにしても, 1 つの値に注目し, その他の値は周辺度数によって決定され, その a, b, c, d によって確率を求め, その確率を累積した値で検定します。

そこで, 次のような $3 * 3$ の頻度表(A:B)のそれぞれの升目の値を $a(+:+)$ として, Fisher の正確検定を応用する方法を考えます。その準備として, どの升目の値も $2 * 2$ の $a(+:+)$ とし, 行の他の値の和を $b(+:-)$ とし, 列の他の値の和を $c(-: +)$ とします。それ以外は $d(-:-)$ です。たとえば A1:B1 のセルならば, $a(+:+): 1, b(+:-) = 2 + 3 = 5, c(-: +) = 4 + 7 = 11, d(-:-) = 5 + 6 + 8 + 9 = 28$ にします。 $a(+:+)$ は A1:B1 の升目の値とするのは当然ですが, $b(+:-) = 2 + 3 = 5$ とするのは, A1(+)かつ B1(-)(A1 であって B1 でない場合)は A1:B2, A1:B3 の 2 つの升目の和と考えられるからです。 $c(-: +)$, $d(-:-)$ も同様です。

A:B	B1	B2	B3
A1	1	2	3
A2	4	5	6
A3	7	8	9

↓

(a, b, c, d)	B1	B2	B3
A1	(1, 5, 11, 28)	(2, 4, 13, 26)	(3, 3, 15, 24)
A2	(4, 11, 8, 22)	(5, 10, 10, 20)	(6, 9, 12, 18)
A3	(7, 17, 5, 16)	(8, 16, 7, 14)	(9, 15, 9, 12)

上表のそれぞれの升目にある 4 つの数字を(a, b, c, d)として, Fisher の正確検定 (上側) をすると, 次の表になります。

Fisher. UP	B1	B2	B3
A1	0.864	0.665	0.456
A2	0.632	0.626	0.624
A3	0.475	0.625	0.749

次の表は L を言語特徴とし, C を年代として想定した相対頻度(R)の分布を示します。

R	C.1	C.2	C.3
L.1	12	10	2
L.2	4	4	6
L.3	1	2	10

この表(R)を，それぞれの升目の(a, b, c, d)を上で説明した方法で計算し，それを使った Fisher の上側正確検定をした結果，次のような上側累積確率 (Fisher.UP)が得られました。

Fisher.UP	C.1	C.2	C.3
L.1	0.018	0.117	1.000
L.2	0.779	0.722	0.352
L.3	0.998	0.969	0.001

上の R の表を見ると，明らかに L1:C1 = 12 が大きな数値(12)を示しているので，それらが有意であること ($p < .05$)は納得できます。UP(L.1:C.2) = .117 よりも UP(L.3:C.3) = .001 のほうが，同じ相対頻度(R) = 10 であるにもかかわらず，小さな確率になって有意性を示していますが ($p < .01$)，この理由は，それぞれの b(+:-), c(-:+)の値が異なるためです。R(L.1:C.2)=10 は

$$b(L.1:C.2) = 12 + 2 = 14$$

$$c(L.1:C.2) = 4 + 2 = 6$$

$$b(L.3:C.3) = 1 + 2 = 3$$

$$c(L.3:C.3) = 2 + 6 = 8$$

なお，上の(a, b, c, d)の表を見てわかるように，すべての升目の和は $a+b+c+d=45$ で同じですが，それぞれの周辺度数($a+b$, $c+d$, $a+c$, $b+d$)は必ずしも一定ではありません。それでもここで Fisher の正確検定の結果を比べているのは，それぞれの a, b, c, d の分布が起こる場合の確率（上側）を比べていることになります。つまり，周辺度数を固定しない片側検定をしています(前掲書：市原(1990: 139)。さらに，ここではとくに検定を目的とせず，有意度の算出のみを目的としているために，多重検定の問題はありません。この方法は $3 * 3$ だけでなく，一般に $n * p$ ($n, p > 1$)の表に適用できます。

■コーパス調査とアンケート調査

正確検定ではそれぞれの横和(e, f)と縦和(g, h)を固定した上で，[a, b, c, d]のすべてのケースを探すために，a の値が選ばれれば，その a の値によって残りの b, c, d の値は自動的に決まります。このように横和・縦和を固定したのは先の紅茶の実験でも同じです。よって正確検定を一般の[a, b, c, d]

のデータに適用するとき、この条件を念頭に入れておかなければなりません。

たとえば4つのスペイン人作家の演劇作品(S)と、メキシコ人作家の演劇(M)で、登場人物の「父親」に向かって普通語 *tú* を使うか(T)、敬語 *usted* を使うか(U)、を調べた、とします。このコーパス調査では正確検定に必要な4つの数値 *a, b, c, d* はそれぞれ S-T, S-U, M-T, M-U の組み合わせになります。そして $[a, b, c, d] = [3, 1, 1, 3]$ という結果になったとします。この場合、それぞれの作品の中の普通語と丁寧語の使用数は固定されているので、正確検定を使うことができます。

$FisherT(3, 1, 1, 3, 2) = .243$ (上側)であることから「スペイン人作家の4作品では父親に向けて *tú* を使い、メキシコ人作家の4作品では *usted* を使う」という結論の有意性を示す度合いは76%ほどです。危険率は非常に高く24.3%です。 $FisherT(4, 0, 0, 4) = .014$ でなければ危険率を5%以下に下げることができません。先の $Fisher(3, 1, 1, 3, 2) = .243$ のそれぞれの頻度を2倍にすると $FisherT(6, 2, 2, 6, 2) = .066$ になり有意度が上昇します。危険率は6.6%にまで下がりました。それぞれの頻度を3倍にすれば $FisherT(9, 3, 3, 9) = .020$ になり、危険率は5%以下になります。このように正確検定を使って信頼できる結果を得るためには調査対象の作品数が8では少なかったと言えるでしょう。

一方、たとえば数人のスペイン人(S)と数人のメキシコ人(M)に「父親に普通語 *tú* を使うか(T)、敬語 *usted* を使うか(U)」を質問する、というアンケート調査を考えると、ここでケース数(S, M)も、TとUの回答数も、自由に変化するはずなので正確検定が使えないことになります。しかし、先の「●Yatesの補正と正確検定」で見たように、周辺度数が自由であっても、正確検定(片側)を使って「スペイン人は父親に向けて *tú* を使い、メキシコ人では *usted* を使う」という仮説の検定をすることができるのです。

このように一般に言語のコーパス調査はそのコーパス(資料)を言語の母集団のサンプルと見なして調査するのではなく、資料そのものを特定の母集団と見なします。もし、コーパスを言語資料の母集団とするのではなく、サンプルだと見なせば、周辺度数は自由になります。また、逆に、アンケートをした2つの集団(S, M)の人数と、その人たちが答えた2つの言語形式(T, U)の数は一定なので、これらの周辺度数(計)を固定して *a, b, c, d* の可能な全変化を想定して確率を計算することも可能です(Fisherの方法)。よって、コーパス調査とアンケート調査のどちらでも、そのデータに限って分析するならばFisherの正確検定が使えます。

■ブラックボックス・リープ・ディスコネクション

書店には統計学の参考書が多く並んでいます。「Excelを使ってこのようにすればよい」と説明する手法の本もたくさんあります。実際に手にとって見ると簡単に統計処理ができるように書かれていて参考になる本もあ

りますが、なかには手法だけを扱って、その応用法についての注意などがなく、数学的な背景については大まかに理解していればよい、という姿勢で書かれているものも多いようです。

たしかに書かれてあるとおりのテクニックを使えばそれなりの結果が出るのですが、どうしてそのような結果が出るのか具体的にわからないことがあります。これでは計算過程が「ブラックボックス」になってしまい、自分が出した結果を説明できません。

参考書の中には説明が「飛躍している」（リープ）と思われるケースもあります。これは説明の段落がどのようにつながるのかわからないような状態です。もしかしたら自分の数学的な知識が不足しているため、リープだと勝手に判断しているのかもしれない。

また、説明の中には「～ということが知られている」「～という公式を使う」というような背景知識に対するリンクになっていることがあります。しかし、私たちが「知られている」という事実や「公式」に疎いとき、背景知識とのリンクは切れてしまっています（ディスコネクション）。

このような理論的な理解がない状態で手法だけを応用してしまうと、結局自分が何をやっているのかわからないのに、自分の名前をつけたレポート・論文・発表を生産してしまうことになります。本人がわかっていないのに、レポート・論文を読む人や発表を聞く人がいるというのは望ましくありません。

そこで、自分にとって、ブラックボックス、リープ、ディスコネクションがあると思われる参考書の説明については、ぜひ自分で実際にいろいろな実験をして納得がいくまで確かめてください。Excel はその実験道具として役立ちます。そして、実験をしながら感覚的に様子がわかったら、今度は統計学や数学の本を読んで数式を理解してください。誰でも難しそうな記号が並んだ数式を目にすると尻込みすることはよくあることですが、そこでしっかり学習してみると案外身近なものであることはよくあることです。理論の理解と実験の順番は逆でも、同時でもよいでしょう。机上の書籍と Excel の往復作業です。いずれにしても自分で納得できた手法を使うことを勧めます。ちょっと面倒かもしれませんが、努力のあとで自分が納得できる成果を得たとき、その達成感が次のステップにつながります。