

## $\chi^2$ 検定

### 1. $\chi^2$ 検定とは

これまでの多変量解析の実習において、2変数の関係については相関分析と回帰分析を学びました。相関分析は、変数  $X$ 、 $Y$  がいずれも量的変数の場合に用いる分析技法でした。回帰分析は、変数  $X$ 、 $Y$  のいずれかが質的変数の場合に用いる分析技法でした。今回学ぶ  $\chi^2$  検定は、扱う変数  $X$ 、 $Y$  がいずれも質的変数であるという点が異なるだけで、分析自体は決して難しいものではありません。

そして、変数  $X$ 、 $Y$  の間にある差が誤差の範囲内なのか、それとも差異として認められるものなのかを統計的に分析する手続きが  $\chi^2$  検定です。具体的にイメージしやすくするために、以下に例に沿って話を進めます。

～例～

国道沿いに2つのガソリンスタンドA店、B店があり、共に給油と洗車のサービスを行っています。石油価格の上昇に伴いガソリンの利幅が減少しているため、洗車によってなんとか利益を確保したいところですが、しかしA店の店長は、「うちは給油をする顧客の数はB店に勝っているが、洗車をする顧客の割合はB店に負けているのではないか？」と思っています。もしこのことが統計的に裏付けられれば、A店の店長は洗車に対する改善策を考えねばならないでしょう。

そこで、A店とB店でサービスを受けた顧客を集計することにしました。その結果が以下の図表1であり、この図表はクロス集計表と呼ばれるものです。このクロス集計表に掲載されている数値を「観測値」と呼びます。

図表1：A店とB店の顧客集計の観測値（単位：人）

	給油	洗車	合計
A店	435	165	600
B店	265	135	400
合計	700	300	1000

ここで便宜上、「A店とB店では洗車をする顧客の割合に差がない」という帰無仮説を立てます。その仮説に基づき、クロス集計表を作成します(図表2参照)。このクロス集計表に掲載されている数値を「期待値」と呼びます。

図表2：A店とB店の顧客集計の期待値(単位：人)

	給油	洗車	合計
A店	420	180	600
B店	280	150	400
合計	700	300	1000

そして、図表1と図表2に掲載された数値の違いを統計的に裏付け、「A店とB店では洗車をする顧客の割合に差があること」そして「A店はB店に比べて洗車をする人の割合が少ない」を実証するための分析技法が $\chi^2$ 検定です。

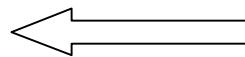
## 2. $\chi^2$ 検定を行うに際して

SASには、 $\chi^2$ 検定用のプロシジャがありません。そのため、度数を算出するためのFREQプロシジャを用い、FREQプロシジャにtable文を設定することによりクロス表を作らせ、そのtable文にchisq(chi square、つまりカイ自乗)オプションをつけることによって、 $\chi^2$ 値とその有意確率値を算出させます。

### インプット例

```
data a; 【 以下に入力するデータを a と名付ける】
input x y w; 【 データ内に変数 x と y と w がある】
cards; 【 ここから次の ; までの間にデータを入力する】
;
proc freq data=a order=; 【 データ a について度数を分析せよ】
tables x*y / chisq norow nocol nopercnt; 【 変数 x,y でクロス表を作成せよ】
weight w; 【 変数 w が度数だ】
output out=b pchi ajchi exact; 【 有意確率等も算出せよ】
proc print data=b; 【 前行の算出結果を出力せよ】
run;
```

1	1	420
1	2	180
2	1	280
2	2	150



アウトプット例

度数	1	2	3	4	5	合計
1	64	43	23	30	40	200
2	72	32	40	66	3	213
合計	136	75	63	96	43	413

x と y の統計量			
統計量	自由度	値	p 値
$\chi^2$ 乗値	4	<u>51.6504</u>	<u>0.001</u>
尤度比 $\chi^2$ 乗値	4	58.0106	0.001
Mantel-haenszel の $\chi^2$ 乗値	1	1.7383	0.1874
ファイ係数		0.3536	
不確実係数		0.3536	
Cramer の V 統計量		0.3536	

3. 報告書の書き方

仮説の提唱

仮説の表記にはいくつかの方法があります。

$X$  と  $Y$  の関係が存在し、その関係があらかじめ仮説化可能な場合、たとえば次のように記述します。

仮説：余暇の過ごし方には男女差がある。

系 1：女性に比べて男性にはボーリングやバッティングセンターを好む人が多い。

系 2：男性に比べて女性にはカラオケやカフェや遊園地を好む人が多い。

$X$  と  $Y$  の関係が存在するが、その内容が予め仮説化不可能な場合、系を省略します。

仮説：余暇の過ごし方には男女差がある。

$X$  と  $Y$  の関係が存在しないと仮説化したい場合、たとえば次のように記述します。

仮説：余暇の過ごし方には男女差がない。

分析の結果

$\chi^2$  値と有意確率 (p 値) を明記してください。有意確率値が小さい値である時 (1%、5%、10%、15% で有意の時)、変数  $X$  と  $Y$  の間は独立 (無関係)。

#### 注意すべきこと

仮説の書き方によって、その結果の書き方が異なるので注意が必要です。

「差がある」という仮説だった場合

有意な場合 差がない 「仮説は支持されなかった(反証された)」

非有意な場合 差がある 「仮説は支持された(正確には反証する証拠はえられなかった)」

ということになります。

「差がない」という仮説だった場合

有意な場合 差がない 「仮説は支持された」

非有意な場合 差がある 「仮説は支持されなかった」

ということになります。

#### 4. 課題

あなたは新規事業を立ち上げる責任者です。あなたが最も興味のある分野に新規事業を立ち上げることができます。その際男女どちらが主要なターゲットとなるか、 $\chi^2$ 検定を用いて統計的に論じてください。なお、どこからのデータを使用しても構いませんが、その際は出所を明らかにしてください。

#### 5. 参考文献・参考 URL

田久浩志,林俊克,小島隆矢(2002),『JMPによる統計解析入門』,オーム社

早稲田大学人間科学部向後千春研究室, <http://kogolab.jp/elearn/hamburger/index.html>

## 「通販における購買意欲の男女差」

第5期 池谷 真剛

### ～課題～

通信販売会社の老舗であるわが社は、海外ブランド G を販売するに先だって、ターゲットを予測したいと考えている。そこで、「通販で海外ブランド G が購入できたら、利用したいと思いますか。」という質問に「1. ぜひ利用したい」、「2. 関心はある」、「3. 利用する気はない」という3つの回答項目を与えて、男女500名に回答を求めた。

回答を集計した結果は下記のとおりであった。

	回答者数	ぜひ利用したい	関心はある	利用する気はない	無回答
全体	500	31.8	32.8	31.4	4.0
男性	240	35.4	31.2	29.6	3.8
女性	260	28.5	34.2	33.1	4.2

### 1. はじめに ～研究目的～

テレビをつけると番組中や、番組と番組の間に、特に深夜の時間帯は日本直販やジャパネットたかたなどによる通信販売の宣伝がよく行われている。通販は、健康器具やジュエリー、衣料品などが主な対象であるが、海外ブランド G を取り扱うとしたら、ターゲットはどこを狙って設定すべきであろうかを考えよう。通販で G を購入できるとしたら男女で購入意欲に差は出るのだろうか。

### 2. 仮説の提唱

予め入手されたデータを基に、独立変数として「性別」を  $X$  とし、従属変数として「通販で海外ブランド G を購入する意欲」を  $Y$  とおき、以下のような仮説を提唱する。

仮説：「 $Y$ : 通販での海外ブランド G の購買意欲」には「 $X$ : 男女差」がある。

### 3. 分析方法の検討、決定

上記の仮説を検定するために  $\chi^2$  検定を用いる。これは、集計値をみて直観的に論ずるのではなく、統計的有意差を検定し論ずるためである。 $\chi^2$  検定を採用した理由としては、独立変数  $X$  が量的な変数である「人数」ではなくて「性別」という質的な変数であり、また、従属変数  $Y$  も「通販での海外ブランド G の購買意欲」という質的な変数であるためである。このように  $X, Y$  ともに量的変数ではなく質的な変数である場合は、 $\chi^2$  検定（カイ自乗検定）という分析方法を採用する。

### 4. 分析結果

上記の仮説を検定するために、入手されたデータに対して  $\chi^2$  検定を行った。モデルの全体的評価に関して、以下の表 1 に要約されるような結果が得られた。モデルに対する検定の結果、 $\chi^2$  値は 1.08 という値を示し、有意確率値  $p$  値は 0.58 で非有意であった。また、表 2 のクロス表より、男女によって購買意欲が異なるという値が得られた。

【表 1】

$\chi^2$ 値	1.08
$p$ 値	0.58

【表 2：クロス集計表（単位：%）】

	ぜひ利用したい	関心はある	利用する気はな
い			
男性	35.4	31.2	29.6

### 5. 考察

分析結果より、 $\chi^2$  検定の有意確率値は非有意であったため、本分析においてこの仮説は支持されたとはいえる。

仮説：「 $Y$ :通販での海外ブランド G の購買意欲」には「 $X$ :男女差」がある 支持された。

男性は、通販での購入に対して 35%が「ぜひ利用したい」で、「利用する気はない」が 30%であった。一方、女性は「ぜひ利用したい」が 28%にとどまり、「利用する気はない」が 33%と高く、男女で逆の数値が得られた。これは、購買行動において男性と女性に違いがあるためだと思われる。労働人口は男性が多くを占めることから考えると、女性に比べて男性は、働いている時間が多くのんびり買い物に出かける時間が少ないと想定される。よって、家でゆっくりしている間にテレビやインターネットで見かけた製品に、

魅了され気に入れば、わざわざ店頭で足を運ぶ面倒が省けるので、買ってしまおうと考えるかもしれない。買い物コストのうち、時間と精神的エネルギーを節約しているのである。一方、女性は時間的余裕があるということもあるが、まず買い物好きというのが一番の要因であると考えられる。店頭で足を運び、ウィンドーショッピングを楽しみつつ、製品を触って素材や容量、コーディネートなども吟味したりする。そして色々な製品を見たり、流行のチェックをしたり、店員との会話を楽しんだりする。おしゃべり好きなこともあって、効果的な人的販売を受けることで気分が高まったところで購入に至るかもしれない。買い物ベネフィットである、買い物の楽しさを重視しているのかもしれない。

#### 6. 終わりに～今後の課題～

今回の分析によって、海外ブランドGの通販導入は女性よりも男性に受容されやすいということが明らかになった。しかし、この結果は海外ブランドGの通販におけるターゲットを明らかにしたまでであり、ビジネスの成否についてはさらなる調査が必要であろう。また、1985年コカ・コーラ社ニュー・コーク導入の失敗例<sup>1</sup>に見られるように、いくら事前の調査が綿密であったとしても、その調査結果がビジネスの成否を正確に予言しうるものではないということにも留意する必要があるかもしれない。

---

<sup>1</sup> 1985年、ペプシコーラに対抗するため、コカ・コーラ社は従来の「コカ・コーラ」に代わって、「ニュー・コーク」を導入した。導入に際しては19万人に及ぶ事前アンケートが実施され、その結果は良好であった。しかし、発売後「ニュー・コーク」に対してアメリカ国民の不満が爆発。コカ・コーラ社は「ニュー・コーク」を廃止し、従来の「コカ・コーラ」を「クラシック・コーク」を復活させることで対応した。