

日本赤十字九州国際看護大学/Japanese Red  
Cross Kyushu International College of  
Nursing

Statistical hypothesis testing

メタデータ	言語: jpn 出版者: 公開日: 2020-08-04 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	<a href="https://jrckicn.repo.nii.ac.jp/records/720">https://jrckicn.repo.nii.ac.jp/records/720</a>

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 International License.



## 第7章 統計的仮説検定

皆さんこんにちは。前回、統計的仮説検定の導入、帰無仮説についてお話ししました。今回は仮説検定の全体像をお話しします。

### 1 帰無仮説による検定の考え方

#### 1) 概要

統計的仮説検定では「2つの変数の間に関連性がある」という仮説を最初から証明しようとせず、その逆に、まず「2つの変数の間に関連性がない」という仮説（帰無仮説）を検討します。関連性を示す証拠として用いるのが検定統計量です。カイ二乗 ( $\chi^2$ ) 値も検定統計量の一つです。変数相互に関連がない場合、帰無仮説が成立している場合、検定統計量は小さな値を取ります。他方、変数相互の関連性が無視できないくらい強くなったとき、「この関連はたまたま、偶然に生じたとするには、あまりにも希少な、小さな確率(例えば 0.05、0.01)で起っている」「偶然とはいい難い」との証拠が固まった時、帰無仮説を棄却（否定）します。では、ずっと「関連性が無い」と言い続け、帰無仮説を保持して来たのに、最後に保持しきれなくなり、帰無仮説を棄却した状態を、どう位置付けたらよいのでしょうか。この場合、何も仮説が無くなったのではなく、帰無仮説と反対の仮説が採択されたと考え、その反対の仮説、つまり「2つの変数の間に関連性がある」という仮説を「対立仮説」といいます。

#### 2) どのようなときに統計的仮説検定を行うか

統計的な仮説検定について教科書は多くのページを割いて述べています。もちろん看護師の国家試験にも出題されます。それくらい大切な考え方ですが、それほどしばしば使うものではありません。例えば今新型コロナウイルスの流行で第2・3波が来ると様々なニュースが報じ、テレビや新聞には毎日、患者数や死亡者数など統計の数字が出てきます。平均値、標準偏差、相関係数などの基本統計量は、現状を知るのに便利です。しかし教科書があれほど重視している仮説検定は、話題になりません。一体どうなっているのでしょうか。どこに仮説検定の話があるか見渡すと、新型コロナウイルスに関連して言えば、治療薬やワクチンがあります。

たとえば、治療薬として既に数か月前から幾つもの名前が上がっていますが、実際に用いられるのは、まだ先の話になりそうです。なぜ時間がかかるのでしょうか。理由の一つは、効果をチェックするための仮説検定を丁寧に行う必要があるからです。「良さそうだから、すぐに使う！」という仮説・行動を取るわけにはいきません。被験者を二群にわけ、投与群と非投与群とで効果を比較する統計的検証、「投与群に効果がある」とする仮説ではなく、「投与群と非投与群とで差が無い」「薬物の投与は疾病の治癒と関連しない」という帰無仮説を検証します。なぜ効果があることを期待しているのに、「効果が無い」という帰無仮説を立てて、慎重に検証を続けるのでしょうか。なぜなら、統計的な検定を行い、帰無仮説を棄却し、効果があると判断すると、簡

単には後戻りできないからです。多くの場合、帰無仮説を棄却すると、その方向で、さまざまな社会的対応が後に続きます。帰無仮説を棄却するというのは、社会的な責任を伴う重い判断だと言えます。

## 2 数表を用いた仮説検定の進め方

検定を行う場合は、その検定統計量の表を読み取ることが必要です。の表は、教科書の末尾にあります。表を読むための基礎知識を以下に示します。

### 1) 主な検定統計量

検定統計量とは、検定の対象となる複数の事象の「関連性や相違の程度」を1つの数値に代表させたもので、与えられた標本のデータから計算されます。

どの課題にどの検定統計量をもちいるか、例を以下に示します。

課題	検定種類	検定統計量	分布と数表
二群で平均を比較	⇒ t 検定	t 値	t 分布
二群でバラツキを比較	⇒ F 検定	F 値	F 分布
三群以上で平均を比較	⇒ F 検定など	F 値	F 分布
期待値と実測値を比較	⇒ $\chi^2$ 検定	$\chi^2$ 値	$\chi^2$ 分布

まず教科書の195頁から数頁をチェックし、数字の並んだ表があるのを確認してください。

### 2) 検定統計量の表の見方

#### (1) 表の構造

197頁のカイ二乗分布表を例に説明します。

表の左端で縦に並ぶ数値、ギリシャ文字でニュー、英語のVの字に似ています、1から30まで、これが自由度です。

表の上端で横に並ぶ数値が有意水準、ギリシャ文字、アルファで表示し、検定統計量に対応する確率を示します。小数点以下の数値が並び、右に行くほど、0.05、0.01など小さくなっています。

そして表の中に並ぶ数値が、検定統計量、カイ二乗値です。

#### (2) カイ二乗検定における表の使い方

- あなたが、観測度数と期待度数から計算したカイ二乗値（4つのセルのカイ二乗値の合計）を用意します。
- 自由度と有意水準を決めます。普通は、自由度=1、有意水準=0.05を用います。
- カイ二乗分布表（197頁）を見て、自由度1、有意水準0.05のカイ二乗値をメモします。（=3.84）
- 上記であなたが計算したカイ二乗値と、表から得た「有意水準に対応するカイ二乗値=3.84」を比較します。この3.84は、帰無仮説の棄却を判断する限界の値ということで、限界値とも呼ばれます。

- あなたが計算したカイ二乗値が 3.84 未満であれば、帰無仮説は棄却できません。「行に示す離散量と列に示す離散量の間には、有意な関連が認められない」と結論します。
- あなたが計算したカイ二乗値が、3.84 以上であれば、帰無仮説は棄却されます。「二つの離散量の間には、有意水準 0.05% で関連性がある」と結論されます。

### 3 もう一步詳しく

・自由度； 自由度とは、さらに詳しくいうと、自由に決められる値という意味で統計的には検定しようという標本の複雑さを表します。カイ二乗検定では表の複雑さに対応し、 $M \times N$  表など複雑なクロス集計表の自由度は  $(m-1)(n-1)$  で計算します。2X2 表はクロス集計表の中では最も単純ですので、自由度は  $(2-1)(2-1) = 1$  となります。

・有意水準； 有意水準はある事象の起こる確率が偶然とは考えにくい、とする判断基準です。統計的検定では、帰無仮説を検討し、どこかの水準で帰無仮説を破棄するという重い判断をする分けですが、その際、この確率が 0.2 (20%)、0.1 (10%) など高めの値だと、帰無仮説が正しいのにそれを棄却してしまう誤り (第一種の過誤) を犯す可能性があります。そこでこの値を普通は 0.05 (5%) さらに慎重に判断する場合は 0.01 (1%) などと設定し、第一種の過誤を少なくすることが一般的です。

### 4 表からコンピューターへ

さて数表は昔から統計計算に欠かせないものとされ、どの統計の教科書にも巻末に数表があり、様々な統計的仮説検定を行う際に活用できます。特にコンピューターの助けを借りることが難しい時代、数表の作成は高度に専門的な作業でした。例えば私が大学院に入った 45 年前は、まだネットもパソコンも存在せず、ソロバン・真空管で動く初歩の計算機・歯車の動きを組み合わせで計算する機械的計算機などが使われていた時代です。自分でカイ二乗分布の曲線を計算することなど、考えられませんでした。

他方、現在はコンピューターが進歩し、実は数表を使わなくても、統計的な数値を直接に計算できます。

たとえば、以前紹介した米国アイオワ大学のサイトを使って、教科書の数表にあるような値を計算してみることができます。

まずパソコンやスマートフォンでアイオワ大学のサイトにアクセスします。このサイトでは様々な確率分布曲線が描けます。

<https://homepage.divms.uiowa.edu/~mbognar/applets/chisq.html>

今、開いているのはカイ二乗分布曲線を描くページです。教科書を持っている人は 72 頁の図をみて下さい。曲線を描くためには、条件の指定が必要です。そこで自由度に対応する空欄 (DF) に 1 を入力すると、赤い矢印で示す曲線が現れました。これが自由度 1 のカイ二乗曲線です。この自由度として 2、3、4 などの数字を入れると、教科書 72 頁のような図が現れます。

今度は、有意水準 P の空欄に 0.05 と入力してください。カイ二乗値が 3.84146 と計算されます。教科書 197 頁の数表にあるカイ二乗値 3.84 と比べると、いま得た数値の方がより正確で桁数が多い

ことが分かります。

最後は、前回の授業で計算したカイ二乗値 16.667 を入力してみてください。P の欄に 0.00004 という数値が現れます。これがカイ二乗曲線から直接に計算した確率、有意確率です。この有意確率が簡単に計算できるなら、実はもう伝統的な数表は必要ありません。有意確率が有意水準より小さな値を取ることが明らかなので、帰無仮説は棄却されます。

## 5 まとめ

以上でカイ二乗検定の基本は終わりです。

統計的な考え方は、帰無仮説の考え方にも現れているように、確率分布を基礎にした様々な数学的な考え方が組み合わせられて出来ています。統計は一度、理論や方法が確定しても、それで終わりではなく、考え方をより洗練させ、より正確な判断が出来るように、検討が続けられています。

今回、動画の中ではお話しませんが、資料の中には、みなさんがよりよく理解できるように、説明を追加しています。余裕があれば、資料にも目を通しておいってください。

## 6 参考

### 1) カイ二乗値について

ピアソンが確立したカイ二乗検定は、離散量から計算されるカイ二乗値を、連続的なカイ二乗分布で近似するため、セルに入る数値が小さいと、近似が不正確になることが指摘されました。

- ・イェーツの補正

これを補正するために、イェーツは 2x2 表の各実測度数と期待度数の差の絶対値から 0.5 を差し引くという簡単な補正法を提案しました。これがイェーツの補正です。

- ・フィッシャーの直接確率

2x2 表のカイ二乗検定で、実測度数が 5 以下、場合によってはゼロなど、とても小さな値を取ると、より厳密な補正方法が必要になります。そこでフィッシャーは、順列組み合わせに基づいたより正確な有意確率の計算方法を提案しました。これがフィッシャーの直接確率法です。

### 2) 自由度 (Degree of freedom)

比較する群に含まれる標本数が多くなると、標本のバラツキが増え、そこに含まれる情報量も増えます。この情報量を表わす目安が自由度です。自由度は検定統計量にも影響を与えます。主要な検定統計量は、自由度別に数字が並んでいます。たとえば 196 頁には平均値の検定に用いる t 分布表が、197 頁にはカイ二乗検定に用いるカイ二乗分布表があります。これらの表では、自由度はギリシャ文字ニューで示されています。英語の V の字に似ていますので、見つけてください。

改めて定義すると、自由度とは「変数のうち独立に (自由に) 選べるものの数」を意味します。たとえば、A さんから D さんまでの 4 人 ( $n=4$ ) について、体重が 46、48、51、53kg とすると平均は 49.5kg、自由に値を取れるデータは 4 人分の体重ですから、自由度は 4、よってデータの個数 ( $n$ ) がそのまま自由度になります。しかし統計学で、平均値から出発し、さらに様々な統計量を計算していく場合、自由度は  $n$  ではなく、 $n-1$  となります。なぜ  $n-1$  になるのか、先ほどの例で言えば、平均値 49.5 という情報を使ってさらに分散などの計算をするとき、平均値

49.5に加えて、AさんからCさんまで3人分の体重の情報があると、4番目のEさんの体重は、情報として必要なくなるからです。（4人分の平均が49.5、AさんからCさんまでが、46、48、51であれば、4人目の体重は既に決定されたこととなります。）このような理由で、統計的仮説検定を行うときには、自由度は $n-1$ を使うのが一般的です。

#### ・クロス集計表の自由度

さて、自由度は $n-1$ と言いましたが、クロス集計表の自由度は、ちょっと独得なので、追加して説明します。皆さんがこれまで集計した2X2表は、クロス集計表の中でも、もっとも単純なもので自由度=1でした。では、より複雑なクロス集計表とはどのようなものでしょうか。また表が複雑だと、自由度はどうなるのでしょうか。

2X2表の場合、行に示した離散量も、列に示した離散量も、それぞれ二つの値（1/0、はい/いいえ、あり/なし）しか取りませんでした。しかし二つ以上の値を取る離散量も多く存在します。たとえば「意思表示；はい/いいえ/どちらでもない」「満足度；とても満足/やや満足/どちらとも言えない/やや不満/とても不満」「回数；0回/1回/2回/3回/・・・n回」などです。こうした離散量をクロス集計する場合は、2X2表では足りず、2XN表、MXN表などが必要になります。MXNのクロス集計表の自由度は $(M-1) \times (N-1)$ となります。

### 3) 有意水準

帰無仮説が成立している状態、行と列に示した二つの離散量が、互いに独立で、両者に何の関連性もない場合、つまり、実測度数が期待度数と一致する場合、カイ二乗値はゼロになります。

他方、実測度数と期待度数の差が大きくなると、カイ二乗値も大きくなります。カイ二乗値がどこまで大きくなったら帰無仮説を棄却するかの基準は有意水準という確率値で示されます。

これが自由度1のカイ二乗曲線です。横軸がカイ二乗値、縦軸はそのカイ二乗値が出現する確率、例えば3は0.3に対応します。この曲線の下面積は合計すると、つまり積分すると1になります。

さて、カイ二乗値が段々に大きくなったとき、どこかで帰無仮説を棄却するかの判断をしなければなりません。たとえばカイ二乗値が1とか2とかで棄却すると、帰無仮説が正しいのに、二つの項目が本当は無関係なのに、その仮説を棄却する間違いを侵す可能性が高いです。棄却域かどうか、その境目のカイ二乗値が棄却限界値（限界値）です。グラフに示した限界値よりも右側の曲線下の領域が棄却域です。統計的な仮説検定では、かなり慎重に棄却域を設定します。曲線下の全面積を1としたとき、棄却域の占める面積の割合を確率、p値で表わし、有意水準と呼びます。有意水準として通常用いられるのは5%、または1%の値です。自由度1のカイ二乗曲線では、有意水準5%のカイ二乗値は3.84、実測値から計算したカイ二乗値がこの値よりも大きいとき、前回の授業中に出てきた例でいえば、カイ二乗値16.667は明らかに3.84より大きいので、有意水準5%で「睡眠時間が6時間未満の場合は、風邪をひきやすい人の割合が、有意に高くなる」などと結論できます。

以上の判断は、有意水準5%ですが、帰無仮説を棄却するかどうかの判断をより厳しくする場合は、有意水準1%、0.1%などを使うこともあり得ます。では様々な有意水準に対応した棄却限界値を知るにはどうしたらよいのでしょうか。また自由度が1より大きい場合はどうしたらよいのでしょうか。コンピューターがあれば、先ほどのアイオワ大学のサイトでのように、直接に棄却限界値などを計算できます。また教科書の最後には、予め計算した表が載っていますので、参照してください。

-----

## 演習問題

1. 動画では前回に続き帰無仮説に触れています。新型コロナウイルス COVID-19 に関して、何かあなたらしい帰無仮説を立ててください。40 字以内で書いてください。
2. 2 X 2 表からの帰無仮説を検定する場合、自由度と有意水準の設定が必要です。有意水準として 0.05 を用いる場合は既に練習しました。では有意水準を 0.01 に変えたとき、検定統計量としてのカイ二乗値は、どのような値になるでしょうか。教科書の表から読み取って、以下に記してください。

3. 以前出てきたのと同じ 2 X 2 表です。

	・アレルギー	
・ペット	あり	なし
いる	12	18
いない	8	62

あなたは既にこの場合のカイ二乗値を、前回の授業で計算しています。有意水準を 0.01 としたとき、この表からの帰無仮説について、あなたはどうか判断しますか。帰無仮説を棄却しますか、それとも維持しますか。あなたの判断とその理由を、以下に 30 字以内で書いてください。

4. 数表や確率分布曲線を作る作業は、以前はとても難しく、数学者が時間をかけておこなっていました。でも今はコンピューターの助けを借りて自分で分布曲線を描けます。以下アイオワ大学のサイトを利用し、自由度に様々な数値（整数）を入力し、曲線を描いてみてください。

<https://homepage.divms.uiowa.edu/~mbognar/applets/chisq.html>

自分で曲線を何本か描いたら教科書 72 頁図 3 - 21 と比較してください。ほぼ同じ？それとも違いがありますか。結果や感想を 40 字以内で書いてください。

## 第8章 調査票の観察と集計

統計学の学習はもう後半に入りました。まず期末レポートの概要を説明します。

「文字数 1000 字以上。図表や数値は文字数に含めない。締切 7 月 22 日。pdf ファイルで提出」  
レポートを書くために必要なデータ、レポートの細かい形式、提出方法などは、次回、第 9 回目の授業で示します。

- ・今回の授業は、レポートを書くための考え方を整理します。
- ・今回の授業では動画は使いません。各設問についている画像の中に、必要な情報を示しています。資料中にも同じ画像があります。

### 課題 1

図に示すのは 3 年前の統計学の時間に、皆さんの先輩が回答した調査票の一部です。ここには連続量あるいは離散量で表せる質問が並んでいます。あなたが関心を持つ項目を、そのデータ形式（連続量か、離散量か）も含めて、三つ挙げてください。

### 課題 2

上記の調査票で、あなたが相関を調べたい項目がありますか。組み合わせを二つ考えてください。回答を入力してください。

### 課題 3

調査票を集計するときは仮説が大切です。もし上記の調査票をあなたが集計するとしたら、どのような仮説を立てますか。まず「原因らしいもの」はどれか、次に「結果らしいもの」はどれかを述べます。また、なぜその二つを選んだのか、理由も記してください。50 文字以内です。

### 課題 4

上記の調査票で調査し、得られた結果の一部（8 名分）を図に示します。8 名分のデータを観察して何か気付いた点があれば、述べてください。このようにデータを観察し見解を述べることは、調分析の出発点として大切です。50 文字以内で書いてください。（新型コロナウイルス COVID-19 が流行している現在の生活と比較すると、気づくことが多いと思います。）

### 課題 5

上記の調査結果から二つの離散量を選んで  $2 \times 2$  表を作成し、行%を計算し、得られた値と考察を 50 文字以内で記してください。

### 課題 6

期末レポートは、データを観察して分かったこと、立てた仮説、仮説に基づいた計算結果、考察など、順序立てて書くことが大切です。その練習として、以上の問（1 から 6 まで）で考えたこと、感じたことをまとめ、分かりやすく 200 文字以内で書いてください。

- 2017年 調査票・保健統計学質問 自分の記号 \_\_\_ \_\_ \_
- A 基本：・性別 ①女 ②男 ・年齢 \_\_\_ 歳 ・身長 \_\_\_ \_\_ cm
- 1 親の職業 ①医療関係 ②医療以外  
 3 生育環境 ①田舎・小都市 ②大都市で育つ  
 4 片づけ・整理 ①得意だ ②苦手だ
- B 現在の生活
- 1 住まい ①一人暮らし ②親元から通う  
 2 パイト ①しない ②する  
 3 大学部活 ①しない ②する  
 4 ボランティア ①しない ②する  
 5 片道通学時間 0---10---20---30---40---50---60---70--  
 6 予習・復習時間 0---10---20---30---40---50---60---70--  
 7 ネット・テレビ時間 0---10---20---30---40---50---60---70--  
 8 パイト時間 0---10---20---30---40---50---60---70--
- C 大学に関連
- 1 講義 ①楽しくない ②楽しい  
 2 実習 ①楽しくない ②楽しい
- D 健康状態
- 1 花粉症等アレルギー ①アレルギーあり ②アレルギーなし  
 2 食べ方 ①食べ過ぎ多い ②食べ過ぎはない  
 3 風邪ひき ①すぐ風邪引く ②風邪引きにくい
- E 生活習慣
- 1 歩くこと ①一日1時間以内 ②一日1時間以上  
 2 食の好き嫌い ①食の好き嫌い多い ②何でもよく食べ

I D	性別	年齢	身長	体重	睡眠時間	A1 親職	A3 生育	A4 片づけ	B1 住まい	B2 バイト	B3 部活	B4 ボランティア	B5 通学時間	B6 予復時間(分)	B7 ネットテレビ時間	B8 バイト時間(分)	C1 講義	C2 実習
1	1	20	153	45	6	2	1	1	2	2	2	2	60	0	200	200	1	1
2	1	19	160	57	6	2	1	1	1	2	1	2	20	60	90	200	1	1
3	1	19	154	53	6	2	1	1	2	2	2	2	20	50	80	200	1	1
4	1	19	153	45	5	2	2	2	2	2	2	2	180	80	60	200	2	1
5	1	19	157	45	7	1	1	2	1	1	1	1	20	120	200	0	1	1
6	2	20	172	53	6	2	1	1	1	2	2	2	10	0	80	200	1	2
7	1	20	153	42	6	2	1	2	1	2	1	2	90	90	90	200	1	1
8	1	20	168	55	5	1	1	2	2	1	2	2	50	30	140	0	1	2