日本赤十字九州国際看護大学/Japanese Red

Cross Kyushu International College of

Nursing

2×2 table and chi-square test

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	https://jrckicn.repo.nii.ac.jp/records/719

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

第６章　クロス集計表とカイ二乗検定

みなさん、こんにちは。今回は２X２表を用いた仮説検定についてお話します。

１　仮説検定の考え方

１）仮説検定とは

まず仮説検定とは何でしょうか。

辞書を引くと「検定」は「一定の基準に基づいて検査し、合格・不合格、等級などを決めること」と定義されています。この検定を統計的な仮説について行うのが「統計的な仮説検定」です。

学生の皆さんは、既に前回の授業で「原因らしい変数⇒結果らしい変数」として仮説を立てています。また２Ｘ２表で行%を計算し、その値から何が考えられるかを考察しています。この考察はとても大切ですが、その一方で、みなさんの主観も含んでいるかもしれません。この主観をできるだけ取り除き、誰もが認める一定の基準（統計的な基準）に基づいて、仮説を受け入れるかどうかを決めることが、統計的な仮説検定です。

２）帰無仮説とは

統計的な仮説検定を行うためには、出発点になる仮説を統計的な発想で、改めて定義する必要があります。これが帰無仮説です。

帰無仮説は「帰」と「無」という二つの文字からいうと、「無に帰すことを前提とした仮説」です。私たちが普通に立てる仮説は「〇が原因で、◇が起こるのではないか？」「△と◇の間には、何か差があるのではないか？」など、観察の結果、事象の相互の関連性や違いに関心を持ったときに、その関連性や違いを具体的に知りたくて立てます。他方、「帰無仮説」というのは「関連性や違い」を否定する、「関連性や違い」は「無である」とする仮説です。

例を挙げると、前回の授業で、みなさんは「睡眠の長さ（６ｈｒ未満／以上）は、風邪への罹患（引きにくい／引きやすい）と関連するかもしれない」「ペットの存在（いる／いない）は、アレルギーへの罹患（あり／なし）と関連するかもしれない」などの仮説を考えました。このように、通常は、差や関連性を疑って仮説を立てます。他方、帰無仮説は「睡眠の長さと、風邪への罹患は、関連しない（両者は独立である）」「ペットの存在と、アレルギーへの罹患は、相互に関連しない（両者は独立である）」となります。

３）帰無仮説を立てる理由

なぜ知りたいことと反対の仮説を、わざわざ立てるのでしょうか。理由としては「両者に関連性がある」「両者に差がある」などの通常の仮説を立てた場合は、小さな関連性から大きな関連性まで、わずかな差から大きな差まで、あらゆる場合について、考えなければならなくなるからです。他方、帰無仮説の場合は「関連性がない」「差がない」という状態だけ検討し、帰無仮説

が否定（棄却）されれば「『関連性がない・差がない』という仮説が棄却された」との判断（検定）を行えるからです。

２　２Ｘ２表における独立性のカイ二乗検定

概要；

カイ二乗検定とは相関係数のときも出てきたピアソンが 1900 年に発表した方法です。２Ｘ２表における帰無仮説は、「二つの変数（行に示す離散量、および列に示す離散量）の間に何の関係もない」「二つの変数は独立である」となります。

検定の手順としては「集計して実際にセルに記入した値（実測度数）」と「帰無仮説による値；独立を仮定した場合に、各セルに期待される値（期待度数）」とを比較し、両者がどれだけ乖離しているか（はなれているか）をカイ二乗値という検定統計量で表わし、カイ二乗値の大きさから、帰無仮説を検定します。

１）２Ｘ２表で実測度数と周辺度数を整理する

ここまでは前回の授業で行っています。思い出してください。正の字を書いて集計し、各セルに書き込んだ４つの値が実測度数です。また行の計、列の計、全体の合計などが周辺度数です。

	B1	B2	計
A1	n_{11}	n_{12}	
A2	n_{21}	n_{22}	
計			

	B1	B2	計
A1			N_{A1}
A2			N_{A2}
計	N_{B1}	N_{B2}	N

	風邪引きやすい	風邪引きにくい	計
睡眠短い	40	20	60
睡眠長い	10	30	40
計	50	50	100

２）２Ｘ２表で期待度数を計算する

帰無仮説が成立している状態、「独立の状態」を数値で表わすのが期待度数です。二つの離散量が独立であれば、その二つの離散量を組み合わせても何の関連性もないわけですから、期待度数は、各離散量を単独で観察した場合の確率を単に掛け算した値で計算できます。

まずワークシートで計算方法を説明します。

計算例；

	風邪引きやすい	風邪引きにくい	計	
睡眠短い	?	?	60	
睡眠長い	?	?	40	
	50	50	100	100.0%

以上の計算を記号で示すと、次のようになります。

	B1	B2	計
A1	$E_{11}=N_{A1}\times N_{B1}/N$	$E_{12}=N_{A1}\times N_{B2}/N$	N_{A1}
A2	$E_{21}=N_{A2}\times N_{B1}/N$	$E_{22}=N_{A2}\times N_{B2}/N$	N_{A2}
計	N_{B1}	N_{B2}	N

ようするに、期待度数の計算には、各セルの実測度数は必要なく、周辺度数の割合さえあれば計算できます。　一般的な式は

セルの期待度数

＝（セルの行の合計度数／全体度数）X（セルの列の合計度数／全体度数）X（全体度数）

＝（セルの行の合計度数）X（セルの列の合計度数）／（全体度数）

	風邪引きやすい	風邪引きにくい	計
睡眠短い	30	30	60
睡眠長い	20	20	40
	50	50	100

３）実測度数と期待度数の差を計算する

すでに実測度数が得られており、新たに期待度数が得られました。実測度数から期待度数を引くと、理論的な独立の状態から現実の数値がどのくらいずれているか、乖離しているかが数値化できます。

	風邪引きやすい	風邪引きにくい	計
睡眠短い	40−30= 10	20−30= −10	
睡眠長い	10−20=−10	30−20=10	

４）カイ二乗値を計算する

さて、実測度数と期待度数のずれはプラスとマイナスの両方があり、どんな２Ｘ２表でも合計するとゼロになってしまいます。そこで、二回目の授業で偏差からバラツキを検討したのと同じ論理に従って、実測度数と期待度数の差を二乗し、すべてプラスの値にします。この二乗した値は、標本数によって大きく異なります。よって標本数による影響を少なくするために、実測度数と期待度数の差の二乗の値を、期待度数で割り算します。こうして得たセルごとの値を足し合わせたものがカイ二乗値です。

各セル；χ^2 (実測度数−期待度数)^2/期待度数

	風邪引きやすい	風邪引きにくい	計
睡眠短い	(40−30)^2/30	(20−30)^2/30	
睡眠長い	(10−20)^2/20	(30−20)^2/20	
計	クロス表全体のχ^2値　16.667		

各セルχ^2 ＝

$$\frac{(\text{実測度数}-\text{期待度数})^2}{\text{期待度数}}$$

$$\chi^2=\sum\frac{(\text{実測度数}-\text{期待度数})^2}{\text{期待度数}}$$

３　カイ二乗検定による判断

さてこれで検定統計量カイ二乗値の値を計算することができました。

実測度数と期待度数との乖離が少なく帰無仮説が成立している場合、すなわち二つの離散量の間に関連性が認められない場合は、カイ二乗値は総体的に小さな値をとります。他方、実測度数と期待度数との乖離が大きくなり、二つの離散量の間に何らかの関連性を否定するのが難しい状況に近づくと、カイ二乗値は総体的に大きな値をとります。

では２Ｘ２表の場合、カイ二乗値がどのくらい以上に大きくなったら、帰無仮説を棄却できるのでしょうか。

一つの目安としては 3.84 が用いられます。これはカイ二乗分布での有意水準 5%自由度 1 における値です。この詳しい意味については次回以降の授業で説明します。

皆さんはこれまで行パーセントからの考察を経験していますが、そこには皆さんの主観が入っていました。

今回、みなさんはカイ二乗値を得たことで、主観ではなく、統計的な客観性に基づいて、帰無仮説を棄却し、「二つの離散量は独立ではない、何らかの関連性がある」という判断を行うことができます。先ほどの例でいえば、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。

４　まとめ

さてこれまで、確率分布分数の考え方、クロス集計表、平均値・分散・標準偏差、共分散・相関係数などを学び、ワークシートで計算演習を重ねてきました。今回の2X2表によるカイ二乗検定は、皆さんが将来何らかの調査を行うときに、最も役立つ検定統計量です。

プリントの最後にはワークシートもありますので是非自分で実測値、期待値、セルごとのカイ二乗値、そして全体のカイ二乗値などを計算してみてください。

なおこれまで授業を重ねるごとに少しずつ複雑な計算を手でしてきましたが、手による計算は今回でほぼ終わりです。これ以降の授業では、手で経験した計算の考え方を、コンピューターを用いて実行することに焦点を移していきます。計算が苦手な人も計算はコンピューターに任せると割り切れば、楽しく学ぶことができると思います。ではまた次回にお会いしましょう。

演習問題

1. 帰無仮説について理解できましたか。あなたは帰無仮説のような考え方をすることがありますか。帰無仮説について、思うことを 40 字以内で書いてください。何を書いても構いません。

2. 前回の授業で出てきたのと同じ２Ｘ２表です。この表について、帰無仮説を立て、40 字以内の文章で示してください。

	・アレルギー	
・ペット	あり	なし
いる	12	18
いない	8	62

3. 上記の２Ｘ２表について、期待度数を計算してください。ワークシートは資料の末尾にあります。答えはセルの順番に、４個の数値で示します。

4. 上記の２Ｘ２表について、各セルのカイ二乗値、および表全体のカイ二乗値を計算してください。答えはセルの順番に数値で示し、最後に表全体のカイ二乗値を示してください。

5. 上記の２Ｘ２表のカイ二乗値から、あなたはどのような結論を出しますか。40 字以内の文章で示してください。

ワークシート：カイ二乗値計算

0 仮説

1 仮説に従って実測度数から２Ｘ２表を作成

原因らしいもの ＼ 結果らしいもの			計
計			

2 行%を計算

			計
			100%
			100%
計			100%

3 期待度数(二変数が独立と仮定)を計算

期待度数＝(そのセルの行の計)×(そのセルの列の計)÷合計

			計
計			

4 実測度数と期待度数の差を計算

			計
計			

5 各セルの χ^2 (実測度数-期待度数)^2/期待度数 を計算

			計
計			

クロス表全体の χ^2 値

6考察

・行%の観察から
　感じること、考えること
・実測と期待の度数を比べ感じること、考えること
・得られた χ^2 値から感じること、考えること

0 仮説

1 仮説に従って実測度数から２×２表を作成

原因らしいもの ＼ 結果らしいもの			計
計			

2 行%を計算

			計
			100%
			100%
計			100%

3 期待度数(二変数が独立と仮定)を計算

期待度数＝(そのセルの行の計)×(そのセルの列の計)÷合計

			計
計			

4 実測度数と期待度数の差を計算

			計
計			

5 各セルの χ^2 (実測度数-期待度数)^2/期待度数 を計算

			計
計			

クロス表全体の χ^2 値

6考察

・行%の観察から
　感じること、考えること
・実測と期待の度数を比べ感じること、考えること
・得られた χ^2 値から感じること、考えること