

1. カテゴリカルデータの分析

仮説検定に関する基本的なことを学んだうえで、今度は心理学でよく使われる検定について見ていきましょう。では、どこから始めればよいのでしょうか。全ての教科書がスタート地点に関する合意を持つわけではないのですが、ここでは“ χ^2 検定”（この章では、“カイ二乗 (にじょう)chi-square” と発音します^{*1}）と“t-検定”（Chapter ??）から始めます。これらの検定は科学的実践において頻繁に使用されており、“回帰”（Chapter ??）や“分散分析”（Chapter ??）ほど強力ではないのですがそれらよりはるかに理解しやすいものとなっています。

“カテゴリカルデータ”という用語は“名義尺度データ”の別名に過ぎません。説明していないのではなく、ただデータ分析の文脈では、“名義尺度データ”よりも“カテゴリカルデータ”という言葉を使う傾向があるのです。なぜかは知りません。なににせよ、**カテゴリカルデータの分析** はあなたのデータが名義尺度の際に適用可能なツールの集合を指示しています。しかし、カテゴリカルデータの分析に使用できるツールには様々なものがあり、本章では一般的なツールの一部のみを取り上げます。

1.1

The χ^2 (カイ二乗) 適合度検定

χ^2 適合度検定は、最も古い仮説検定の一つです。この検定は世紀の変わり目に Karl Pearson 氏が考案したもので (Pearson1900)、Ronald Fisher 氏によっていくつかの修正が加えられました (Fisher1922)。名義尺度変数に関する観測度数分布が期待度数分布と合致するかどうかを調べます。例えば、ある患者グループが実験的処置を受けており、彼・彼女らの状況が改善されたか、変化がないか、悪化したかを確認するために健康状態が評価されたとします。各カテゴリー（改善、変化なし、悪化）の数値が、標準的な処置条件で期待される数値と一致するかどうかを判断するために、適

^{*1}また“カイ二乗 (にじょう)chi-squared”とも呼ばれる

合性検定は適用できます。もう少し、心理学を交えて考えてみましょう。

1.1.1 カードデータ

何年にもわたる多くの研究が、人が完全にランダムにふるまおうとすることの難しさを示しています。ランダムに「行動」しようとしても、我々はパターンや構造に基づいて考えてしまいます。そのため、「ランダムになにかをしてください」と言われたとしても人々が実際に行うことはランダムなものにはなりません。結果として、人のランダム性 (あるいは非ランダム性) に関する研究は、我々が世界をどのように捉えているのかについての深遠な心理学的問いを数多く投げかけます。このことを念頭に置いて、非常に簡単な研究について考えてみましょう。シャッフルされたカードのデッキを想像して、このデッキの中から「ランダムに」一枚のカードを頭の中で選ぶようお願いしたとします。一枚目のカードを選んだ後、二枚目のカードを心の中で選択してもらいます。二つの選択に関して、注目するのは選ばれたカードのマーク (ハート、クラブ、スペード、ダイヤモンド) です。これをたとえば $N = 200$ にやってもらうよう依頼した後、選択されようとしたカードが本当にランダムに選ばれているかどうかをデータを確認して調べてみましょう。データは `randomness.csv` に入っており、JASP で開くと 3 つの変数が表示されるでしょう。変数 `id` は各参加者に対する一意識別子であり、二つの変数 `choice_1` と `choice_2` は参加者が選択したカードのマークを意味しています。

今回は、参加者の選んだ最初の選択肢に注目してみましょう。‘Descriptives’ - ‘Descriptive Statistics’ の下にある [Frequency tables](#) オプションを選択して、選択された各マークの数をカウントしてみましょう。以下が得られたものです:

クラブ	ダイヤモンド	ハート	スペード
35	51	64	50

この小さな度数分布表はとても有益です。この表を見れば、人はクラブよりもハートを選びやすいかもしれないというわずかなヒントを得られますが、それが実際にそうであるのか偶然の賜物であるのかどうかは見るだけでは明らかではありません。なので、それを知るためにはなんらかの統計分析をしなければならないでしょう。それが、次のセクションでお話しすることになります。

よろしい。ここからは、先ほどの表を分析対象のデータとして扱います。しかしながら、このデータについて数学的に語らなければならないために、表記の意味について明確にしておくことは大事でしょう。数学的表記では、人が読める単語である "observed (観測された)" を文字 O に短縮して、観測位置を示すために下付き文字を使用します。なので、この表における二番目の観測変数は数学では O_2 として記述します。日本語表記と数学記号の関係を以下に示します:

ラベル	インデックス, i	数学, シンボル	数値
クラブ, ♣	1	O_1	35
ダイヤモンド, ◇	2	O_2	51
ハート, ♥	3	O_3	64
スペード, ♠	4	O_4	50

これではっきりしたでしょう。また、数学者は特定の事柄よりも一般的な事柄について話したがるので、 O_i という表記が見られるでしょう。これは、 i 番目のカテゴリーに属する観測変数を意味します (i は 1、2、3、4 のいずれか)。最後に、観測された頻度数に言及したい場合、統計家は観測値をベクトル^{*2}に分類します。これは、太字を使用して \mathbf{bmO} とします。

$$\mathbf{O} = (O_1, O_2, O_3, O_4)$$

繰り返しますが、これは新しいものでも興味深いものでもありません。ただの表記です。 $\mathbf{O} = (35, 51, 64, 50)$ ということ、私がしているのは観測された度数表の記述 (i.e., **observed**) ですが、数学表記を用いてそれを参照します。

1.1.2 帰無仮説と対立仮説

先ほどのセクションで指摘したように、我々の研究仮説は「人はカードをランダムに選択しない」です。これから行いたいことはこれを統計的仮説に変換してから、それらの仮説に関する統計検定を構築することです。説明予定のテストは**ピアソンの χ^2 (カイ二乗) 適合度検定**であり、よくあることですが、まずは帰無仮説の注意深い構築から始めなければなりません。今回はかなり簡単です。まず、帰無仮説を言葉にしてみましょう：

H_0 : 4 つ全てのマークは同じ確率で選択される

さて、これは統計学なので、同じことを数学っぽく言えなければなりません。これをするために、 j 番目のマークが選ばれる場合の真の確率を参照するときには表記 P_j を用いましょう。もし帰無仮説が真であれば、4 つのマークがそれぞれ 25% の確率で選択されます。言い換えれば、帰無仮説は $P_1 = .25, P_2 = .25, P_3 = .25$ そして $P_4 = .25$ としたものです。ただし、観測された頻度数をデータ全体の要約ベクトル \mathbf{O} として分類するように、帰無仮説と対応する確率として \mathbf{P} を用います。そのため、帰無仮説を記述する確率の集合を $\mathbf{P} = (P_1, P_2, P_3, P_4)$ とすると、以下のようになります：

H_0 : $\mathbf{P} = (.25, .25, .25, .25)$

この例では、帰無仮説は全ての確率が互いに等しい確率のベクトル \mathbf{P} と対応します。しかし、常に

^{*2}ベクトルは同じ基本型のデータ要素のシーケンスです

そうである必要はありません。例えば、もし実験課題で他のマークの 2 倍クラブが含まれているデッキを想像してもらう場合には、帰無仮説は $P = (.4, .2, .2, .2)$ となるでしょう。確率がすべて正の値であり、その合計が 1 である限りは、それは帰無仮説として正当な選択です。ですが、適合度検定では一般的に全てのカテゴリーが同様の確率である帰無仮説を用います。そのため、ここではそれに固執します。

対立仮説 H_1 はどうでしょうか？ 我々の関心は、関係する確率が全て同じではないこと（つまり、人々の選択が完全にランダムではなかったこと）を実証することです。その結果、「人にやさしい（負担の小さい）」バージョンの仮説はこんな感じです：

H_0 : 4 つのマークが同じ確率で選択される

H_1 : 少なくとも 1 つのマークの選択確率が 0.25 ではない

そして「数学者にやさしい」バージョンはこうです：

H_0 : $P = (.25, .25, .25, .25)$

H_1 : $P \neq (.25, .25, .25, .25)$

1.1.3 “適合度” 検定の統計量

この段階で、観測された頻度数 O と検定予定の帰無仮説と対応する確率の集合である P を我々は手にしています。ここでしたいことは、帰無仮説検定の構築です。いつものように、 H_1 に対して H_0 を検定したい場合には、検定統計量が必要です。適合度検定の基本的なトリックは、データが帰無仮説にどれだけ「近いのか」を測定する検定統計量を組み立てることです。もし帰無仮説が真であるときの「期待値」がデータと似ていなければ、帰無仮説は真ではないでしょう。オーケー、帰無仮説が真であるならどうなるだろう？ 正しい言い方をすれば、「期待度数」とは何かということです。 $N = 200$ の観測データがあり、（もし帰無仮説が真であれば）ハートの選択確率は $P_3 = .25$ で、ハートの期待値は $200 \times .25 = 50$ ですね？ より具体的には、もし E_i が、「帰無仮説が真であるときにカテゴリー i の選択数の期待値」とすると次のようになります。

$$E_i = N \times P_i$$

この計算はとても簡単です。もし 4 つのカテゴリーに分類されうる 200 個の観測データがあつて、カテゴリー全ての選択確率が同じだとすれば、各カテゴリーの観測データは 50 であると期待されますよね？

さて、これをどのように検定統計量に変換するのでしょうか？ 明らかに、我々のしたいことは各カテゴリーの期待値 (E_i) と観測値 (O_i) の比較です。そしてこの比較に基づいて、我々は良い検定統計量を導き出すことができるはずです。まずは、帰無仮説が期待した結果と我々が実際に得られた結果との差を計算しましょう。つまり、「観測値から期待値を引いた」差得点である $O_i - E_i$ を計算

します。これを図示すると次の表のようになります。

		♣	◇	♡	♠
期待度数	E_i	50	50	50	50
観測度数	O_i	35	51	64	50
差得点	$O_i - E_i$	-15	1	14	0

つまり我々の計算によって、帰無仮説の予測よりも人はハートを多く、クラブを少なく選択していることがわかりました。しかし、ちょっと考えてみると、この素朴な違いは、私たちが求めているものとはちょっと違うようです。直観的に、帰無仮説の予測が少なすぎた場合（ハート）も予測が多すぎた場合（クラブ）も同程度によくないことに感じられます。なので、クラブではマイナスでハートではプラスだというのはちょっと変な感じです。これを解決する一つの簡単な方法は全てを二乗することで、ここでは二乗された差を計算します $(O_i - E_i)^2$ 。前回同様、これは手作業でできます：

		♣	◇	♡	♠
期待度数	E_i	50	50	50	50
観測度数	O_i	35	51	64	50
差得点	$O_i - E_i$	-15	1	14	0
二乗差	$(O_i - E_i)^2$	225	1	196	0

さあ、これで一步前進です。今手にしているものは、帰無仮説が悪い予測をしたときには大きく（クラブとハート）、良い予測をしたときには小さくなる数の集合です。次は、後述予定の技術的理由により、それらの数を期待度数 E_i で割って、**調整された 二乗値** $\frac{(E_i - O_i)^2}{E_i}$ を計算しています。今回の例では全てのカテゴリで $E_i = 50$ となるので、あまり面白い計算ではないですが、とりあえずやってみましょう：

		♣	◇	♡	♠
期待度数	E_i	50	50	50	50
観測度数	O_i	35	51	64	50
差得点	$O_i - E_i$	-15	1	14	0
二乗差	$(O_i - E_i)^2$	225	1	196	0
調整済み二乗差	$(O_i - E_i)^2 / E_i$	4.50	0.02	3.92	0.00

要するに、ここで得たのは4つの「エラー」得点で、それぞれが観測度数に対する帰無仮説の予測から生じた「間違い」の大きさを示しています。そして、これを一つの有用な検定統計量に変換するためには、それらの数を単に足し合わせることが一つのやり方です。その結果を **適合度:goodness-of-fit** とよび、慣習的には χ^2 (カイ二乗) または (頭文字をとって) GOF とよばれています。 $4.50 + 0.02 + 3.92 + 0.00 = 8.44$ として計算可能です。

k をカテゴリーの総数だとすれば (i.e., カードデータの例だと $k = 4$)、 χ^2 統計量は以下のように与えられます：

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

直観的に、もし χ^2 が小さければ観測データ O_i が帰無仮説の予測する E_i に近づき、帰無仮説を棄却するためには大きな χ^2 が必要となるはずです。

計算の結果、カードデータセットでは $\chi^2 = 8.44$ という値が得られました。次の疑問はこれが帰無仮説を棄却するのに十分な値なのか、ということです。

1.1.4 適合度の標本分布

χ^2 の値が帰無仮説を棄却するほどに大きいかどうかを決めるために、帰無仮説が真である場合の χ^2 に関する標本分布はどうなるかを理解する必要があります。ということで、今回のセクションではそれをやっていきます。この標本分布がどのように構成されるかの詳細をお見せして、次のセクションではそれを仮説検定の構築に用います。さて本題。標本分布が $k - 1$ の自由度を持つ χ^2 (カイ二乗) 分布であると喜んで受け入れる人は、このセクションの残りをスキップできます。しかし、もし適合度検定の仕組みを理解したいひとは、ぜひこの先をお読みください。

よし、帰無仮説が実際に真であると仮定しましょう。もしそうであれば、ある観測変数が i 番目のカテゴリーに属する真の確率は P_i となります。まあ結局、それはほぼほぼ帰無仮説の定義です。これが意味するところについて考えます。これは、「自然」が、重み付きのコイン (i.e., 表が出る確率は P_i) を裏返すことで観測値がカテゴリー i に含まれるかどうかを決定するというようなものです。したがって、自然がこれらのコインの N を反転させること (データセット内の観測ごとに 1 つ) を想像して、正確に O_i を頭に浮かべることで、観測された頻度 O_i を考えることができます。明らかに、これは実験を考える上ではかなり奇妙なやり方です。ですが、このシナリオには見覚えがありますね (と私は期待してますよ)。セクション ??、二項分布のケースとまったく同じ設定です。言い換えれば、帰無仮説が真であれば、観測された度数は二項分布のサンプリングによって生成されたことになります。

$$O_i \sim \text{Binomial}(P_i, N)$$

中心極限定理 (Section ??) の説明を思い出すと、特に N が大きく P_i が 0 または 1 に近すぎない時に、二項分布は正規分布と近似して見えるようになります。いいかえれば、 $N \times P_i$ が十分に大きければいいのです。また、別の言い方をすれば、期待度数 E_i が十分に大きい場合 O_i の理論的な分布は近似的に正規分布となります。さらにいえば、 O_i が正規分布していれば、そのとき $(O_i - E_i)/\sqrt{E_i}$ も正規分布します。 E_i は固定の値なので、 E_i を引いて $\sqrt{E_i}$ で割ることで正規分布の平均と標準偏差が変化しますが、それだけです。では早速、適合度統計量とはなにかについて見ていきましょう。