第 6 章：カテゴリデータの推測

OpenIntro Statistics 第 4 版（日本語版）

原著スライド：Mine Çetinkaya-Rundel（OpenIntro）

2 人の科学者が、ある薬が高血圧に有効かどうかを調べたいと考えています。1 人目の科学者は、高血圧を持つ 1000 人にその薬を投与し、血圧が下がった人数を記録しようとしています。2 人目の科学者は、高血圧を持つ 500 人に薬を投与し、別の高血圧を持つ 500 人には投与しないで、両グループで血圧が下がった人数を比較しようとしています。この薬を試験するためにより良い方法はどちらですか？

(a) 1000 人全員に薬を投与する
(b) 500 人に投与し、500 人には投与しない

2 人の科学者が、ある薬が高血圧に有効かどうかを調べたいと考えています。1 人目の科学者は、高血圧を持つ 1000 人にその薬を投与し、血圧が下がった人数を記録しようとしています。2 人目の科学者は、高血圧を持つ 500 人に薬を投与し、別の高血圧を持つ 500 人には投与しないで、両グループで血圧が下がった人数を比較しようとしています。この薬を試験するためにより良い方法はどちらですか？

(a) 1000 人全員に薬を投与する

(b) *500 人に投与し、500 人には投与しない*

GSS の結果

GSS は同じ質問を行っており、以下は 2010 年の調査の回答の分布です：

1000 人全員に薬を投与する	99
500 人に投与し、500 人には投与しない	571
合計	670

母数（パラメータ）と点推定値

実験計画について正しい直感を持つ、つまり「500 人に投与し、500 人には投与しない」と答えるアメリカ人全体の割合を推定したいと思います。関心のある母数（パラメータ）と点推定値は何ですか？

- 関心のある母数（パラメータ）：全アメリカ人のうち、実験計画について正しい直感を持つ人の割合。

$$p \ \left(\text{母集団の割合}\right)$$

- 点推定値：標本抽出されたアメリカ人のうち、実験計画について正しい直感を持つ人の割合。

$$\hat{p} \ \left(\text{標本の割合}\right)$$

割合の推測

実験計画について正しい直感を持つ、つまり「500 人に投与し、500 人には投与しない」と答えるアメリカ人全体の割合は何パーセントですか？

- この研究上の問いは信頼区間を使って答えることができます。信頼区間は常に次の形をとります：

$$\text{点推定値} \pm \text{誤差の範囲}$$

- また、誤差の範囲 = 臨界値 × 点推定値の標準誤差 であることも知っています。

標本割合の標準誤差

$$SE_{\hat{p}} = \sqrt{\frac{p\,(1-p)}{n}}$$

標本割合もほぼ正規分布に従う

割合の中心極限定理

標本割合は、母集団の割合 p を平均とし、標準誤差 $\sqrt{\frac{p\ (1-p)}{n}}$ を持つ正規分布にほぼ従います。

$$\hat{p} \sim N\left(mean = p, SE = \sqrt{\frac{p\ (1-p)}{n}}\right)$$

もちろん、これはある条件のもとでのみ成り立ちます…

何か予想できますか？

独立な観測値、少なくとも *10* の成功と *10* の失敗

注：p が未知の場合（ほとんどのケース）、標準誤差の計算に $\hat{p}$ を使用します。

実験計画の話に戻ろう…

GSS によると、670 人中 571 人（85%）のアメリカ人が実験計画に関する質問に正しく回答しました。実験計画について正しい直感を持つアメリカ人全体の割合を、95%信頼区間を用いて推定してください。

与えられた値：$n = 670, \hat{p} = 0.85$。まず条件を確認します。

1. 独立性：標本は無作為に抽出されており、670 < 全アメリカ人の 10%であるため、ある回答者の回答が別の回答者の回答と独立していると仮定できます。
2. 成功・失敗条件：571 人が正しく回答（成功）し、99 人が誤って回答（失敗）しており、両方とも 10 以上です。

$n = 670, \hat{p} = 0.85$ が与えられており、標本割合の標準誤差は $SE = \sqrt{\frac{p(1-p)}{n}}$ であることを学びました。95%信頼区間の正しい計算式はどれですか？

(a) $0.85 \pm 1.96 \times \sqrt{\frac{0.85 \times 0.15}{670}}$

(b) $0.85 \pm 1.65 \times \sqrt{\frac{0.85 \times 0.15}{670}}$

(c) $0.85 \pm 1.96 \times \frac{0.85 \times 0.15}{\sqrt{670}}$

(d) $571 \pm 1.96 \times \sqrt{\frac{571 \times 99}{670}}$

$n = 670, \hat{p} = 0.85$ が与えられており、標本割合の標準誤差は $SE = \sqrt{\frac{p(1-p)}{n}}$ であることを学びました。95%信頼区間の正しい計算式はどれですか？

(a) $0.85 \pm 1.96 \times \sqrt{\frac{0.85 \times 0.15}{670}} \rightarrow (0.82, 0.88)$

(b) $0.85 \pm 1.65 \times \sqrt{\frac{0.85 \times 0.15}{670}}$

(c) $0.85 \pm 1.96 \times \frac{0.85 \times 0.15}{\sqrt{670}}$

(d) $571 \pm 1.96 \times \sqrt{\frac{571 \times 99}{670}}$

標本サイズの選択

95%信頼区間の誤差の範囲を 1%に抑えるためには、何人を標本抽出する必要がありますか？

$$ME = z^{\star} \times SE$$

$$
\begin{aligned}
0.01 &\geq 1.96 \times \sqrt{\frac{0.85 \times 0.15}{n}} \rightarrow \text{前の研究の } \hat{p} \text{ を使用} \\
0.01^2 &\geq 1.96^2 \times \frac{0.85 \times 0.15}{n} \\
n &\geq \frac{1.96^2 \times 0.85 \times 0.15}{0.01^2} \\
n &\geq 4898.04 \rightarrow n \text{ は少なくとも 4,899 必要}
\end{aligned}
$$

過去の研究がない場合は？

$\cdots\ \hat{p} = 0.5$ を使用します

なぜですか？

- 何もわからない場合、50-50 が最も合理的な推測です
- $\hat{p} = 0.5$ は最も保守的な推定値を与えます — 最大の標本サイズが必要

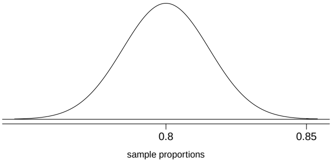
割合の CI と HT

- 成功・失敗条件：
  - CI：少なくとも 10 の観測された成功と失敗
  - HT：帰無値を使って計算した、少なくとも 10 の期待される成功と失敗
- 標準誤差：
  - CI：観測された標本割合を使って計算：$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
  - HT：帰無値を使って計算：$SE = \sqrt{\frac{p_0(1-p_0)}{n}}$

GSS によると、670 人中 571 人（85%）のアメリカ人が実験計画に関する質問に正しく回答しました。このデータは、80%以上のアメリカ人が実験計画について正しい直感を持つという説得力のある証拠を提供しますか？

$$H_0 : p = 0.80 \qquad H_A : p > 0.80$$

$$\begin{aligned} SE &= \sqrt{\frac{0.80 \times 0.20}{670}} = 0.0154 \\ Z &= \frac{0.85 - 0.80}{0.0154} = 3.25 \\ p\text{値} &= 1 - 0.9994 = 0.0006 \end{aligned}$$



p 値が低いため、H_0 を棄却します。このデータは、80%以上のアメリカ人が実験計画について正しい直感を持つという説得力のある証拠を提供します。

2006 年のギャラップ調査に回答した 1,001 人のアメリカ人のうち、11%が宗教的な理由からハロウィンのお祝いに反対していると回答しました。95%信頼水準で、この調査の誤差の範囲は ±3% です。この研究結果に関するニュース記事には「10%以上のアメリカ人が宗教的な理由からハロウィンのお祝いに反対している」と書かれています。95%信頼水準で、このニュース記事の主張は正当化されますか？

(a) はい
(b) いいえ
(c) 判断できない

2006 年のギャラップ調査に回答した 1,001 人のアメリカ人のうち、11%が宗教的な理由からハロウィンのお祝いに反対していると回答しました。95%信頼水準で、この調査の誤差の範囲は $\pm 3\%$ です。この研究結果に関するニュース記事には「10%以上のアメリカ人が宗教的な理由からハロウィンのお祝いに反対している」と書かれています。95%信頼水準で、このニュース記事の主張は正当化されますか？

(a) はい
(b) いいえ
(c) 判断できない

まとめ - 1つの割合の推測

- 母集団のパラメータ：p、点推定値：$\hat{p}$
- 条件：
  - 独立性
    - 無作為標本かつ 10%条件
  - 少なくとも 10 の成功と失敗
    - 満たさない場合 $\rightarrow$ 無作為化
- 標準誤差：$SE = \sqrt{\frac{p(1-p)}{n}}$
  - CI の場合：$\hat{p}$ を使用
  - HT の場合：p_0 を使用

北極の氷冠の融解

科学者たちは、地球温暖化が今後 100 年以内に極地域に大きな影響を与える可能性があると予測しています。その一つとして、北極の氷冠が完全に融解する可能性があります。もしそれが実際に起きた場合、あなたはどの程度気になりますか？ 非常に気になる、ある程度気になる、少し気になる、まったく気にならない？

(a) 非常に気になる
(b) ある程度気になる
(c) 少し気になる
(d) まったく気にならない

GSS の結果

GSS は同じ質問を行っており、以下は 2010 年の GSS 調査およびデューク大学の入門統計学の学生グループからの回答の分布です：

	GSS	デューク大
非常に気になる	454	69
ある程度気になる	124	30
少し気になる	52	4
まったく気にならない	50	2
合計	680	105

母数（パラメータ）と点推定値

- 関心のある母数（パラメータ）：全デューク大学生と全アメリカ人のうち、北極の氷冠の完全融解によって非常に気になると答える割合の差。

$$p_{\text{デューク}} - p_{\text{米国}}$$

- 点推定値：標本抽出されたデューク大学生と標本抽出されたアメリカ人のうち、北極の氷冠の完全融解によって非常に気になると答える割合の差。

$$\hat{p}_{\text{デューク}} - \hat{p}_{\text{米国}}$$

割合を比較するための推測

- 詳細は以前と同じです…
- CI：点推定値 ± 誤差の範囲
- HT：$Z = \frac{\text{点推定値} - \text{帰無値}}{SE}$ を使って適切な p 値を求めます。
- 点推定値の適切な標準誤差（$SE_{\hat{p}_{\text{デューク}} - \hat{p}_{\text{米国}}}$）が唯一の新しい概念です。

2 つの標本割合の差の標準誤差

$$SE_{(\hat{p}_1 - \hat{p}_2)} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

割合の差の CI の条件

1. グループ内の独立性：
   - 米国グループは無作為に標本抽出されており、デューク大学グループも無作為標本を代表していると仮定します。
   - 105 < 全デューク大学生の 10%、680 < 全アメリカ人の 10% です。

   標本内のデューク大学生の態度は互いに独立しており、標本内の米国住民の態度も互いに独立していると仮定できます。
2. グループ間の独立性：標本抽出されたデューク大学生と米国住民は互いに独立しています。
3. 成功・失敗条件：
   2 つのグループで少なくとも 10 の観測された成功と 10 の観測された失敗があること。

北極の氷冠の融解によって非常に気になると答えるデューク大学生とアメリカ人の割合の差 $(p_{\text{デューク}} - p_{\text{米国}})$ について、95%信頼区間を構築してください。

データ	デューク大	米国
非常に気になる	69	454
それ以外	36	226
合計	105	680
$\hat{p}$	0.657	0.668

$$
\begin{aligned}
& (\hat{p}_D - \hat{p}_{US}) \pm z^\star \times \sqrt{\frac{\hat{p}_D(1-\hat{p}_D)}{n_D} + \frac{\hat{p}_{US}(1-\hat{p}_{US})}{n_{US}}} \\
= \quad & (0.657 - 0.668) \pm 1.96 \times \sqrt{\frac{0.657\times0.343}{105} + \frac{0.668\times0.332}{680}} \\
= \quad & -0.011 \pm 1.96 \times 0.0497 \\
= \quad & -0.011 \pm 0.097 \\
= \quad & (-0.108, 0.086)
\end{aligned}
$$

北極の氷冠の融解によって非常に気になると答えるデューク大学生全体の割合が、同様に答えるアメリカ人全体の割合と異なるかどうかを検定するための正しい仮説の組み合わせはどれですか？

(a) $H_0 : p_{\text{デューク}} = p_{\text{米国}}$
$H_A : p_{\text{デューク}} \neq p_{\text{米国}}$

(b) $H_0 : \hat{p}_{\text{デューク}} = \hat{p}_{\text{米国}}$
$H_A : \hat{p}_{\text{デューク}} \neq \hat{p}_{\text{米国}}$

(c) $H_0 : p_{\text{デューク}} - p_{\text{米国}} = 0$
$H_A : p_{\text{デューク}} - p_{\text{米国}} \neq 0$

(d) $H_0 : p_{\text{デューク}} = p_{\text{米国}}$
$H_A : p_{\text{デューク}} < p_{\text{米国}}$

(a) と (c) の両方が正しいです。

北極の氷冠の融解によって非常に気になると答えるデューク大学生全体の割合が、同様に答えるアメリカ人全体の割合と異なるかどうかを検定するための正しい仮説の組み合わせはどれですか？

(a) $H_0 : p_{\text{デューク}} = p_{\text{米国}}$
$H_A : p_{\text{デューク}} \neq p_{\text{米国}}$

(b) $H_0 : \hat{p}_{\text{デューク}} = \hat{p}_{\text{米国}}$
$H_A : \hat{p}_{\text{デューク}} \neq \hat{p}_{\text{米国}}$

(c) $H_0 : p_{\text{デューク}} - p_{\text{米国}} = 0$
$H_A : p_{\text{デューク}} - p_{\text{米国}} \neq 0$

(d) $H_0 : p_{\text{デューク}} = p_{\text{米国}}$
$H_A : p_{\text{デューク}} < p_{\text{米国}}$

(a) と *(c)* の両方が正しいです。

1 つの割合の検定への振り返り

- 母集団の割合の信頼区間を構築する際は、観測された成功数と失敗数が少なくとも 10 であるか確認します。

$$n\hat{p} \geq 10 \qquad n(1-\hat{p}) \geq 10$$

- 母集団の割合の仮説検定を行う際は、期待される成功数と失敗数が少なくとも 10 であるか確認します。

$$np_0 \geq 10 \qquad n(1-p_0) \geq 10$$

割合のプールされた推定値

- $H_0 : p_1 = p_2$ の場合に 2 つの割合を比較する際、各標本の期待される成功数と失敗数を計算するために使える与えられた帰無値がありません。
- そのため、まず 2 つのグループに共通の（プールされた）割合を求め、それを分析に使用する必要があります。
- これは単純に、観測総数のうちの成功の総数の割合を求めることを意味します。

割合のプールされた推定値

$$\hat{p} = \frac{\text{成功数}_1 + \text{成功数}_2}{n_1 + n_2}$$

北極の氷冠の融解によって非常に気になると答えるデューク大学生とアメリカ人のプールされた割合を計算してください。プールされた推定値はどちらの標本割合（$\hat{p}_{デューク}$ または $\hat{p}_{米国}$）に近いですか？ なぜですか？

データ	デューク大	米国
非常に気になる	69	454
それ以外	36	226
合計	105	680
$\hat{p}$	0.657	0.668

$$\begin{aligned}\hat{p} &= \frac{\text{成功数}_1 + \text{成功数}_2}{n_1 + n_2} \\ &= \frac{69 + 454}{105 + 680} = \frac{523}{785} = 0.666\end{aligned}$$

これらのデータは、北極の氷冠の融解によって非常に気になると答えるデューク大学生全体の割合が、同様に答えるアメリカ人全体の割合と異なることを示唆していますか？ 検定統計量、p 値を計算し、データの文脈で結論を解釈してください。

データ	デューク大	米国
非常に気になる	69	454
それ以外	36	226
合計	105	680
$\hat{p}$	0.657	0.668

$$
\begin{aligned}
Z &= \frac{(\hat{p}_{\text{デューク}} - \hat{p}_{\text{米国}})}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_{\text{デューク}}} + \frac{\hat{p}(1-\hat{p})}{n_{\text{米国}}}}} \\
&= \frac{(0.657 - 0.668)}{\sqrt{\frac{0.666 \times 0.334}{105} + \frac{0.666 \times 0.334}{680}}} = \frac{-0.011}{0.0495} = -0.22 \\
p\text{値} &= 2 \times P(Z < -0.22) = 2 \times 0.41 = 0.82
\end{aligned}
$$

まとめ - 2つの割合の比較

- 母集団のパラメータ：$(p_1 - p_2)$、点推定値：$(\hat{p}_1 - \hat{p}_2)$
- 条件：
  - グループ内の独立性
    - 両グループで無作為標本かつ 10%条件を満たす
  - グループ間の独立性
  - 各グループで少なくとも 10 の成功と失敗
    - 満たさない場合 → 無作為化（6.4 節）
- $SE_{(\hat{p}_1 - \hat{p}_2)} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$
  - CI の場合：$\hat{p}_1$ と $\hat{p}_2$ を使用
  - HT の場合：
    - $H_0 : p_1 = p_2$ の場合：$\hat{p}_{pool} = \frac{\text{成功}_1 + \text{成功}_2}{n_1 + n_2}$ を使用
    - $H_0 : p_1 - p_2 =$（0 以外の値）の場合：$\hat{p}_1$ と $\hat{p}_2$ を使用
      - これはかなり稀

参考 - 標準誤差の計算

	1 標本	2 標本
平均	$SE = \frac{s}{\sqrt{n}}$	$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
割合	$SE = \sqrt{\frac{p(1-p)}{n}}$	$SE = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

- 平均を扱う場合、σ が既知であることは非常に稀なので、通常 s を使用します。
- 割合を扱う場合、
  - 仮説検定を行う場合、p は帰無仮説から来ます
  - 信頼区間を構築する場合、代わりに $\hat{p}$ を使用します

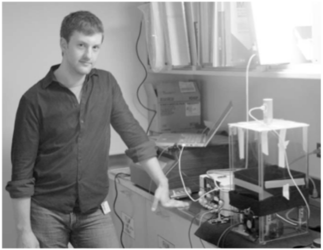
ウェルドンのサイコロ

- ウォルター・フランク・ラファエル・ウェルドン（1860–1906）は、イギリスの進化生物学者で、生物測定学の創設者の一人です。フランシス・ゴールトンおよびカール・ピアソンとともに、Biometrika の共同創設編集者でした。
- 1894 年に、彼は 12 個のサイコロを 26,306 回振り、5 か 6 が出た回数（成功とみなした）を記録しました。
  - 5 か 6 が期待よりも頻繁に出ることが観察され、ピアソンはこれがサイコロの構造によるものだと仮説を立てました。安価なサイコロの多くは凹んだ点を持ち、反対の面の合計が 7 になるため、6 の目のある面は 1 の目のある反対の面より軽くなります。

ラビーのサイコロ

- 2009 年、ザカライア・ラビー（シカゴ大学）は、手製のサイコロ投げ・点数カウント機械を使ってウェルドンの実験を再現しました。

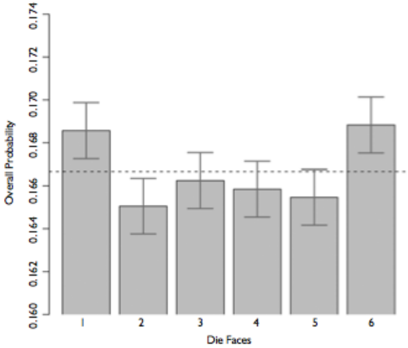
  http://www.youtube.com/watch?v=95EErdouO2w



- 転がして画像処理するのに約 20 秒かかりました。
  - 1 日あたり約 150 枚の画像を手動で処理する必要がありました。
  - このペースで、ウェルドンの実験は 6 日以上かけて再現されました。
  - 推薦読書：*http://galton.uchicago.edu/about/docs/labby09dice.pdf*

ラビーのサイコロ（続き）

- ラビーは実際には、ウェルドンが観察した現象（5 と 6 の頻度が高い）を観察しませんでした。
- 自動化によりラビーは 1894 年のウェルドンよりも多くのデータを収集でき、「成功」と「失敗」を記録する代わりに、各サイコロの個別の目の数を記録しました。



期待度数

ラビーは 12 個のサイコロを 26,306 回振りました。各面が等しい確率で出るとすれば、1、2、⋯、6 はそれぞれ何回出ることが期待されますか？

(a) $\frac{1}{6}$
(b) $\frac{12}{6}$
(c) $\frac{26{,}306}{6}$
(d) $\frac{12 \times 26{,}306}{6}$

期待度数

ラビーは 12 個のサイコロを 26,306 回振りました。各面が等しい確率で出るとすれば、1、2、⋯、6 はそれぞれ何回出ることが期待されますか？

(a) $\frac{1}{6}$
(b) $\frac{12}{6}$
(c) $\frac{26{,}306}{6}$
(d) $\frac{12\times 26{,}306}{6} \quad = 52,612$

ラビーの結果のまとめ

以下の表は、ラビーの実験から得られた観測度数と期待度数を示しています。

結果	観測度数	期待度数
1	53,222	52,612
2	52,118	52,612
3	52,465	52,612
4	52,338	52,612
5	52,244	52,612
6	53,285	52,612
合計	315,672	315,672

期待度数がすべての結果で同じなのに、観測度数が異なるのはなぜですか？ 一見して、観測度数と期待度数の間に不一致があるように見えますか？

仮説の設定

これらのデータは、観測度数と期待度数の間の不一致の説得力のある証拠を提供しますか？

H_0: 観測度数と期待度数の間に不一致はありません。観測度数は期待度数と同じ分布に従います。

H_A: 観測度数と期待度数の間に不一致があります。観測度数は期待度数と同じ分布に従いません。サイコロを振ったときに出る面にバイアスがあります。

仮説の評価

- これらの仮説を評価するために、観測度数が期待度数からどれだけ異なるかを定量化します。
- 標本変動（偶然）のみによって期待される値から大きく逸脱している場合、対立仮説の強い証拠となります。
- これは適合度検定と呼ばれ、観測データが期待される分布にどの程度適合しているかを評価します。

検定統計量の構造

- 検定統計量の一般的な形は

$$\frac{\text{点推定値} - \text{帰無値}}{\text{点推定値の SE}}$$

- この構造は以下の 2 つに基づいています：
  1. 帰無仮説が真の場合の期待値と点推定値の差を特定し、
  2. その差を点推定値の標準誤差で標準化する。

  これらの 2 つの考え方が、度数データの適切な検定統計量の構築に役立ちます。

カイ二乗統計量

度数を扱い、観測度数が期待度数からどれだけ離れているかを調べる場合、カイ二乗（χ^2）統計量と呼ばれる新しい検定統計量を使用します。

χ^2 統計量

$$\chi^2 = \sum_{i=1}^{k} \frac{(O-E)^2}{E} \qquad \text{ここで } k = \text{セルの総数}$$

カイ二乗統計量の計算

結果	観測度数	期待度数	$\frac{(O-E)^2}{E}$
1	53,222	52,612	$\frac{(53{,}222-52{,}612)^2}{52{,}612} = 7.07$
2	52,118	52,612	$\frac{(52{,}118-52{,}612)^2}{52{,}612} = 4.64$
3	52,465	52,612	$\frac{(52{,}465-52{,}612)^2}{52{,}612} = 0.41$
4	52,338	52,612	$\frac{(52{,}338-52{,}612)^2}{52{,}612} = 1.43$
5	52,244	52,612	$\frac{(52{,}244-52{,}612)^2}{52{,}612} = 2.57$
6	53,285	52,612	$\frac{(53{,}285-52{,}612)^2}{52{,}612} = 8.61$
合計	315,672	315,672	24.73

なぜ二乗するのか？

観測された結果と期待された結果の差を二乗することは 2 つのことをします：

- 二乗された標準化された差はすべて正になります。
- すでに異常に見えた差は、二乗後にさらに大きくなります。

以前にこのようなことを見たことがありますか？

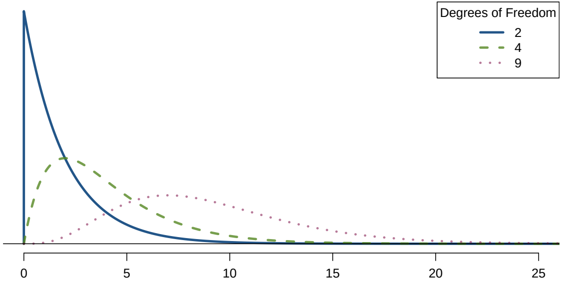
カイ二乗分布

- 計算した χ^2 統計量が異常に高いかどうかを判断するために、まずその分布を説明する必要があります。
- カイ二乗分布は自由度（*df*）と呼ばれる 1 つのパラメータのみを持ち、これが分布の形状、中心、広がりに影響します。

復習：これまでに *3* つの他の連続分布を見てきました：

- 正規分布：平均と標準偏差の *2* つのパラメータを持つ単峰で対称な分布
- *t* 分布：自由度の *1* つのパラメータを持つ単峰で対称な分布
- *F* 分布：分子の自由度（グループ間分散）と分母の自由度（グループ内分散）の *2* つのパラメータを持つ単峰で右に歪んだ分布

次のうち誤りはどれですか？



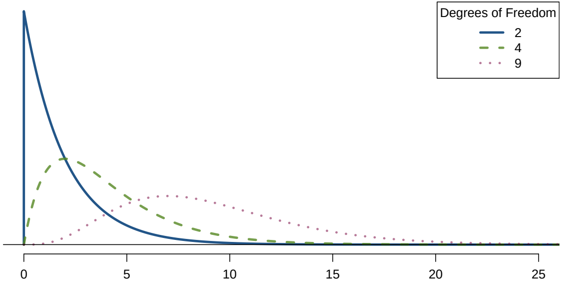
df が増加するにつれて、

(a) χ^2 分布の中心も増加する

(b) χ^2 分布のばらつきも増加する

(c) χ^2 分布の形状がより歪んでくる（正規分布に近くなくなる）

次のうち誤りはどれですか？



df が増加するにつれて、

(a) χ^2 分布の中心も増加する

(b) χ^2 分布のばらつきも増加する

(c) χ^2 分布の形状がより歪んでくる（正規分布に近くなくなる）

カイ二乗曲線の下の面積を求める

- p 値 = カイ二乗分布の裾の面積（通常通り）
- このためにテクノロジーやカイ二乗確率表を使用できます。

カイ二乗曲線の下の面積を求める（続き）

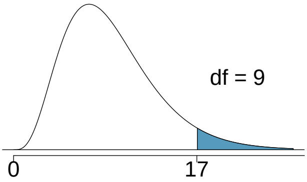
$df = 6$ の χ^2 曲線で、カットオフ値 10 より上の影付き部分の面積を推定してください。

```
> pchisq(q = 10, df = 6, lower.tail = FALSE)

[1] 0.124652
```

カイ二乗曲線の下の面積を求める（続き）

$df = 9$ の χ^2 曲線で、カットオフ値 17 より上の影付き部分の面積を推定してください。



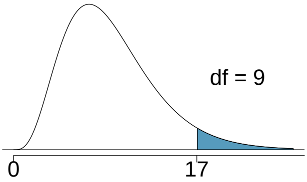
(a) 0.05
(b) 0.02
(c) 0.02 と 0.05 の間
(d) 0.05 と 0.1 の間
(e) 0.01 と 0.02 の間

```
> pchisq(q = 17, df = 9, lower.tail = FALSE)

[1] 0.04871598
```

カイ二乗曲線の下の面積を求める（続き）

$df = 9$ の χ^2 曲線で、カットオフ値 17 より上の影付き部分の面積を推定してください。



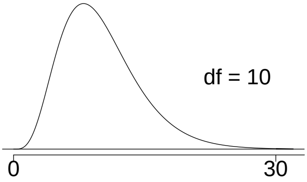
(a) 0.05
(b) 0.02
(c) *0.02 と 0.05 の間*
(d) 0.05 と 0.1 の間
(e) 0.01 と 0.02 の間

```
> pchisq(q = 17, df = 9, lower.tail = FALSE)

[1] 0.04871598
```

カイ二乗曲線の下の面積を求める（もう一問）

$df = 10$ の χ^2 曲線で、30 より上の影付き部分の面積を推定してください。



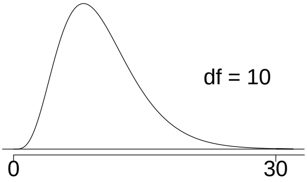
(a) 0.3 より大きい
(b) 0.005 と 0.001 の間
(c) 0.001 未満
(d) 0.001 より大きい
(e) この表では判断できない

```
> pchisq(q = 30, df = 10, lower.tail = FALSE)

[1] 0.0008566412
```

カイ二乗曲線の下の面積を求める（もう一問）

$df = 10$ の χ^2 曲線で、30 より上の影付き部分の面積を推定してください。



(a) 0.3 より大きい
(b) 0.005 と 0.001 の間
(c) *0.001 未満*
(d) 0.001 より大きい
(e) この表では判断できない

```
> pchisq(q = 30, df = 10, lower.tail = FALSE)

[1] 0.0008566412
```

ラビーのサイコロに戻ろう

- 研究上の問いは：これらのデータは、観測度数と期待度数の間の不一致の説得力のある証拠を提供しますか？
- 仮説は：
  - H_0: 観測度数と期待度数の間に不一致はありません。観測度数は期待度数と同じ分布に従います。
  - H_A: 観測度数と期待度数の間に不一致があります。観測度数は期待度数と同じ分布に従いません。サイコロを振ったときに出る面にバイアスがあります。
- 検定統計量 $\chi^2 = 24.67$ を計算しました。
- df があれば、裾の面積（p 値）を計算して仮説に関する決定を下すことができます。

適合度検定の自由度

- 観測データが期待される分布にどの程度適合しているかを評価する適合度検定では、自由度はセル数（k）から 1 を引いた値として計算されます。

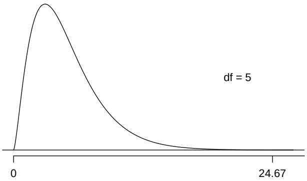
$$df = k - 1$$

- サイコロの結果では $k = 6$ なので、

$$df = 6 - 1 = 5$$

カイ二乗検定の p 値の求め方

カイ二乗検定の p 値は、計算された検定統計量より上の裾の面積として定義されます。



p 値 $= P(\chi^2_{df=5} > 24.67)$ は 0.001 未満

仮説検定の結論

p 値が 0.001 未満と計算されました。有意水準 5%で、仮説検定の結論は何ですか？

(a) H_0 を棄却する：データはサイコロが公正であるという説得力のある証拠を提供します。

(b) H_0 を棄却する：データはサイコロに偏りがあるという説得力のある証拠を提供します。

(c) H_0 を棄却しない：データはサイコロが公正であるという説得力のある証拠を提供します。

(d) H_0 を棄却しない：データはサイコロに偏りがあるという説得力のある証拠を提供します。

仮説検定の結論

p 値が 0.001 未満と計算されました。有意水準 5%で、仮説検定の結論は何ですか？

(a) H_0 を棄却する：データはサイコロが公正であるという説得力のある証拠を提供します。

(b) H_0 を棄却する：データはサイコロに偏りがあるという説得力のある証拠を提供します。

(c) H_0 を棄却しない：データはサイコロが公正であるという説得力のある証拠を提供します。

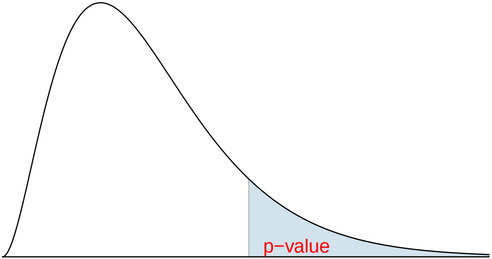
(d) H_0 を棄却しない：データはサイコロに偏りがあるという説得力のある証拠を提供します。

実は…

- 1-6 の軸は他の 2 つの軸（2-5 と 3-4）より一貫して短く、1 の目と 6 の目のある面が他の面より大きいという仮説を支持しています。
- 凹んだ点のために 5 と 6 がより頻繁に出るというピアソンの主張は、これらのデータによって支持されません。
- カジノで使用されるサイコロはフラッシュフェースを持ち、点は周囲の素材と同じ密度のプラスチックで埋められ、精密にバランスが取られています。

まとめ：カイ二乗検定の p 値

- カイ二乗検定の p 値は、計算された検定統計量の上の裾の面積として定義されます。
- 検定統計量は常に正であり、高い検定統計量ほど帰無仮説からの強い逸脱を意味するからです。



カイ二乗検定の条件

1. 独立性：表に度数を寄与する各ケースは、表の他のすべてのケースと独立していなければなりません。
2. 標本サイズ：各特定のシナリオ（すなわち、セル）は少なくとも 5 つの期待されるケースがなければなりません。
3. $df > 1$：自由度は 1 より大きくなければなりません。

条件の確認を怠ると、意図せず検定の誤り率に影響を与える可能性があります。

2009 年イラン選挙

2009 年のイラン選挙では選挙不正が多く話題になりました。ここでは、選挙前に実施された世論調査のデータ（観測データ）と選挙で報告された得票率を比較して、両者が同じ分布に従うかどうかを確認します。

	世論調査での観測	選挙での報告
候補者	有権者数	得票率
(1) アフマディネジャド	338	63.29%
(2) ムサビ	136	34.10%
(3) その他の候補者	30	2.61%
合計	504	100%
	↓	↓
	観測度数	期待される分布

仮説

報告された得票率と世論調査の分布が異なるかどうかを検定するための仮説は何ですか？

H_0: 世論調査から観測された度数は、報告された得票率と同じ分布に従います。

H_A: 世論調査から観測された度数は、報告された得票率と同じ分布に従いません。

検定統計量の計算

候補者	観測数（世論調査）	報告得票率（選挙）	期待数（世論調査）
(1) アフマディネジャド	338	63.29%	504 × 0.6329 = 319
(2) ムサビ	136	34.10%	504 × 0.3410 = 172
(3) その他の候補者	30	2.61%	504 × 0.0261 = 13
合計	504	100%	504

$$
\begin{aligned}
\frac{(338-319)^2}{319} &= 1.13 \\
\frac{(136-172)^2}{172} &= 7.53 \\
\frac{(30-13)^2}{13} &= 22.23 \\
\chi^2_{df=2} &= 30.89
\end{aligned}
$$

結論

これらの計算に基づいて、仮説検定の結論は何ですか？

(a) p 値が低く、H_0 は棄却されます。世論調査から観測された度数は、報告された得票率と同じ分布に従いません。

(b) p 値が高く、H_0 は棄却されません。世論調査から観測された度数は、報告された得票率と同じ分布に従います。

(c) p 値が低く、H_0 は棄却されます。世論調査から観測された度数は、報告された得票率と同じ分布に従います。

(d) p 値が低く、H_0 は棄却されません。世論調査から観測された度数は、報告された得票率と同じ分布に従いません。

結論

これらの計算に基づいて、仮説検定の結論は何ですか？

(a) p 値が低く、H_0 は棄却されます。世論調査から観測された度数は、報告された得票率と同じ分布に従いません。

(b) p 値が高く、H_0 は棄却されません。世論調査から観測された度数は、報告された得票率と同じ分布に従います。

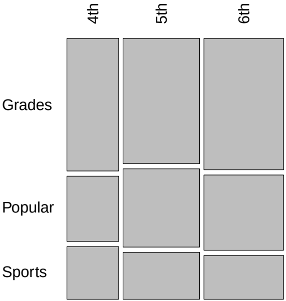
(c) p 値が低く、H_0 は棄却されます。世論調査から観測された度数は、報告された得票率と同じ分布に従います。

(d) p 値が低く、H_0 は棄却されません。世論調査から観測された度数は、報告された得票率と同じ分布に従いません。

人気のある子供たち

popular データセットでは、4〜6 年生が、良い成績・運動能力・人気のどれが最も重要かを尋ねられました。学年と最も重要な要因で分けた二元表を示します。目標が学年によって異なることを示す証拠はありますか？

	成績	人気	スポーツ
4 年生	63	31	25
5 年生	88	55	33
6 年生	96	55	32



カイ二乗独立性検定

- 仮説は：

  H_0: 学年と目標は独立しています。目標は学年によって異なりません。

  H_A: 学年と目標は従属しています。目標は学年によって異なります。

- 検定統計量は次のように計算されます：

$$\chi^2_{df} = \sum_{i=1}^{k} \frac{(O-E)^2}{E} \qquad \text{ここで} \qquad df = (R-1)\times(C-1),$$

ここで k はセル数、R は行数、C は列数です。

注：一元表と二元表では df の計算が異なります。

- p 値は χ^2_{df} 曲線の、計算された検定統計量より上の面積です。

二元表の期待度数

二元表の期待度数

$$\text{期待度数} = \frac{(\text{行合計}) \times (\text{列合計})}{\text{表の合計}}$$

	成績	人気	スポーツ	合計
4 年生	*63*	*31*	25	119
5 年生	88	55	33	176
6 年生	96	55	32	183
合計	247	141	90	478

$$E_{\text{行 } 1,\ \text{列 } 1} = \frac{119 \times 247}{478} = 61 \qquad E_{\text{行 } 1,\ \text{列 } 2} = \frac{119 \times 141}{478} = 35$$

二元表の期待度数

ハイライトされたセルの期待度数は何ですか？

	成績	人気	スポーツ	合計
4 年生	63	31	25	119
5 年生	88	*55*	33	176
6 年生	96	55	32	183
合計	247	141	90	478

(a) $\frac{176\times141}{478}$

(b) $\frac{119\times141}{478}$

(c) $\frac{176\times247}{478}$

(d) $\frac{176\times478}{478}$

$\rightarrow$ *52*

期待より多くの 5 年生が人気を目標としている

二元表の期待度数

ハイライトされたセルの期待度数は何ですか？

	成績	人気	スポーツ	合計
4 年生	63	31	25	119
5 年生	88	55	33	176
6 年生	96	55	32	183
合計	247	141	90	478

(a) $\frac{176 \times 141}{478}$

(b) $\frac{119 \times 141}{478}$

(c) $\frac{176 \times 247}{478}$

(d) $\frac{176 \times 478}{478}$

$\rightarrow$ *52*

期待より多くの 5 年生が
人気を目標としている

二元表での検定統計量の計算

期待度数は観測度数の隣に青色で示されています。

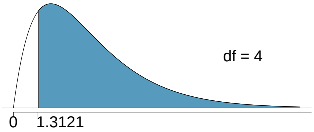
	成績	人気	スポーツ	合計
4年生	63 *61*	31 *35*	25 *23*	119
5年生	88 *91*	55 *52*	33 *33*	176
6年生	96 *95*	55 *54*	32 *34*	183
合計	247	141	90	478

$$\chi^2 = \sum \frac{(63-61)^2}{61} + \frac{(31-35)^2}{35} + \cdots + \frac{(32-34)^2}{34} = 1.3121$$
$$df = (R-1) \times (C-1) = (3-1) \times (3-1) = 4$$

p 値の計算

この仮説検定の正しい p 値はどれですか？

$$\chi^2 = 1.3121 \qquad df = 4$$

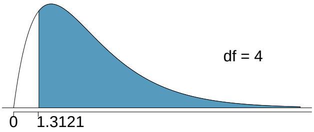


(a) 0.3 より大きい
(b) 0.3 と 0.2 の間
(c) 0.2 と 0.1 の間
(d) 0.1 と 0.05 の間
(e) 0.001 未満

p 値の計算

この仮説検定の正しい p 値はどれですか？

$$\chi^2 = 1.3121 \qquad df = 4$$



(a) *0.3 より大きい*
(b) 0.3 と 0.2 の間
(c) 0.2 と 0.1 の間
(d) 0.1 と 0.05 の間
(e) 0.001 未満

結論

これらのデータは、目標が学年によって異なることを示す証拠を提供しますか？

H_0: 学年と目標は独立しています。目標は学年によって異なりません。

H_A: 学年と目標は従属しています。目標は学年によって異なります。

p 値が高いため、H_0 を棄却しません。データは学年と目標が従属しているという説得力のある証拠を提供しません。目標は学年によって異ならないようです。