

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：

点推定値 ± 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。

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

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？

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。

- 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。
- 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？

- $0.85 \pm 1.96 \times \sqrt{\frac{0.85 \times 0.15}{670}} \rightarrow (0.82, 0.88)$
- $0.85 \pm 1.65 \times \sqrt{\frac{0.85 \times 0.15}{670}}$
- $0.85 \pm 1.96 \times \frac{0.85 \times 0.15}{\sqrt{670}}$
- $571 \pm 1.96 \times \sqrt{\frac{571 \times 99}{670}}$

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？

$$ME = z^* \times SE$$

$$0.01 \geq 1.96 \times \sqrt{\frac{0.85 \times 0.15}{n}} \rightarrow \text{前の研究の } \hat{p} \text{ を使用}$$

$$0.01^2 \geq 1.96^2 \times \frac{0.85 \times 0.15}{n}$$

$$n \geq \frac{1.96^2 \times 0.85 \times 0.15}{0.01^2}$$

$$n \geq 4898.04 \rightarrow n \text{ は少なくとも } 4,899 \text{ 必要}$$

過去の研究がない場合は？

… $\hat{p} = 0.5$ 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： $SE = \sqrt{\frac{p_0(1-p_0)}{n}}$

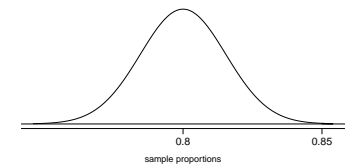
GSSによると、670人中571人（85%）のアメリカ人が実験計画に関する質問に正しく回答しました。このデータは、80%以上のアメリカ人が実験計画について正しい直感を持つという説得力のある証拠を提供しますか？

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

$$SE = \sqrt{\frac{0.80 \times 0.20}{670}} = 0.0154$$

$$Z = \frac{0.85 - 0.80}{0.0154} = 3.25$$

$$p\text{値} = 1 - 0.9994 = 0.0006$$



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

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 → 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

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

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。

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

- 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。

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

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（ $\hat{p}_{\text{デューク}}$ または $\hat{p}_{\text{米国}}$ ）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

カイ二乗曲線の下での面積を求める

- 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。

カイ二乗曲線の下での面積を求める (続き)

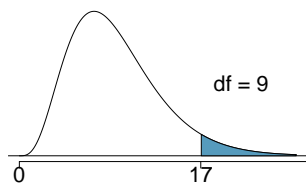
$df = 6$ の χ^2 曲線で、カットオフ値 10 より上の影付き部分の面積を推定してください。

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

```
[1] 0.124652
```

カイ二乗曲線の下での面積を求める (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。



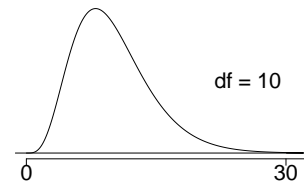
- (a) 0.05
- (b) 0.02
- (c) 0.02 と 0.05 の間
- (d) 0.05 と 0.1 の間
- (e) 0.01 と 0.02 の間

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

```
[1] 0.04871598
```

カイ二乗曲線の下での面積を求める (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
```

ラビース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。

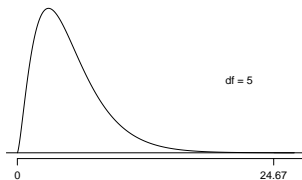
$$df = k - 1$$

- 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 = $P(\chi_{df=5}^2 > 24.67)$
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？

- 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。
- 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。
- 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。
- 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。

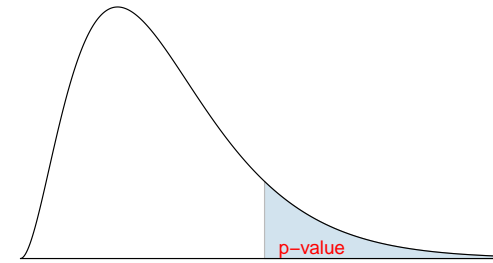
実は…

- 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。

候補者	世論調査での観測 有権者数	選挙での報告 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？

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%	$504 \times 0.6329 = 319$
(2) ムサビ	136	34.10%	$504 \times 0.3410 = 172$
(3) その他の候補者	30	2.61%	$504 \times 0.0261 = 13$
合計	504	100%	504

$$\frac{(338-319)^2}{319} = 1.13$$

$$\frac{(136-172)^2}{172} = 7.53$$

$$\frac{(30-13)^2}{13} = 22.23$$

$$\chi^2_{df=2} = 30.89$$

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がありますか？

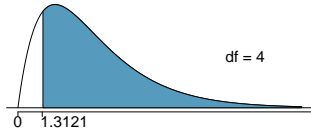
	成績	人気	スポーツ
4年生	63	31	25
5年生	88	55	33
6年生	96	55	32

	成績	人気	スポーツ
Grades			
Popular			
Sports			

p値の計算

この仮説検定の正しいp値はどれですか？

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



- (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。