確率と統計　第12回

仮説検定: それは事実か思い込みか

7月6日

Table of contents

§1 仮説検定の枠組み

§2 信頼区間を用いた仮説検定

§3 P 値を使ったより汎用的な仮説検定

ファクトフルネスからの例題

教育をうけた人たちがどれだけ世界の実情を把握しているかを知るためにファクトフルネスの著者たちは以下のような問いについての正答率を調査した.

ファクトフルネスからの例題

教育をうけた人たちがどれだけ世界の実情を把握しているかを知るためにファクトフルネスの著者たちは以下のような問いについての正答率を調査した.

問. 世界で何らかの病気に対して予防接種を受けている 1 歳児はどのくらいいるか.

(a) 20%
(b) 50%
(c) 80%

ファクトフルネスからの例題

教育をうけた人たちがどれだけ世界の実情を把握しているかを知るためにファクトフルネスの著者たちは以下のような問いについての正答率を調査した.

問. 世界で何らかの病気に対して予防接種を受けている 1 歳児はどのくらいいるか.

(a) 20%
(b) 50%
(c) 80%

(正解は 80%)

帰無仮説と対立仮説

さて大卒者の正答率はあてずっぽうよりも高いだろうか低いだろうか．それを知るためには以下の 2 つの相反する仮説のどちらが正しいかを検証すればいい．

帰無仮説と対立仮説

さて大卒者の正答率はあてずっぽうよりも高いだろうか低いだろうか．それを知るためには以下の 2 つの相反する仮説のどちらが正しいかを検証すればいい．

> H_0: 正答率は $p = 0.33$ である．(あてずっぽうと変わらない)

帰無仮説と対立仮説

さて大卒者の正答率はあてずっぽうよりも高いだろうか低いだろうか．それを知るためには以下の 2 つの相反する仮説のどちらが正しいかを検証すればいい．

H_0: 正答率は $p = 0.33$ である．(あてずっぽうと変わらない)

H_A: 正答率は $p \neq 0.33$ である．(あてずっぽうより高いか低い)

帰無仮説と対立仮説

一般に統計学では, 検証したい主張 A があるとき, まずはそれと反対の主張を用意する.

帰無仮説と対立仮説

一般に統計学では, 検証したい主張 A があるとき, まずはそれと反対の主張を用意する.
そして主張 A を支持するための強い証拠がある場合に, その反対の主張を棄却して主張 A を採用する.

帰無仮説と対立仮説

一般に統計学では, 検証したい主張Aがあるとき, まずはそれと反対の主張を用意する.
そして主張Aを支持するための強い証拠がある場合に, その反対の主張を棄却して主張Aを採用する.
このときの主張Aを対立仮説といい, Aと反対の主張を帰無仮説という.

帰無仮説と対立仮説

一般に統計学では, 検証したい主張 A があるとき, まずはそれと反対の主張を用意する.
そして主張 A を支持するための強い証拠がある場合に, その反対の主張を棄却して主張 A を採用する.
このときの主張 A を対立仮説といい, A と反対の主張を帰無仮説という.
つまり対立仮説を支持する強い証拠によって帰無仮説が棄却されることで初めて対立仮説を採用するという回りくどい方法をとる.

帰無仮説と対立仮説

先ほどの例では大卒者が世界の実情について知識を持っているかどうかを検証したいので,

帰無仮説と対立仮説

先ほどの例では大卒者が世界の実情について知識を持っているかどうかを検証したいので,

- H_0: 正答率は $p = 0.33$ である. (あてずっぽうと変わらない)
- H_A: 正答率は $p \neq 0.33$ である. (あてずっぽうより高いか低い)

帰無仮説と対立仮説

先ほどの例では大卒者が世界の実情について知識を持っているかどうかを検証したいので,

- H_0: 正答率は $p = 0.33$ である. (あてずっぽうと変わらない)
- H_A: 正答率は $p \neq 0.33$ である. (あてずっぽうより高いか低い)

のうち, H_0 が帰無仮説で H_A が対立仮説となる.

例題: どちらが帰無/対立仮説か

ある被告が有罪か無罪かを判断するための裁判を行う．このときの帰無仮説 H_0 と対立仮説 H_A はそれぞれ何か.

例題: どちらが帰無/対立仮説か

ある被告が有罪か無罪かを判断するための裁判を行う．このときの帰無仮説 H_0 と対立仮説 H_A はそれぞれ何か.

H_0: 被告は無罪である.

H_A: 被告は有罪である.

とすると, 十分な証拠がある場合にのみ有罪判決が下る一般的な裁判になる. (H_0 と H_A が逆だと...)

例題: どちらが帰無/対立仮説か

ある新薬の有効性を調べたい. このときの帰無仮説と対立仮説はそれぞれ何か.

例題: どちらが帰無/対立仮説か

ある新薬の有効性を調べたい. このときの帰無仮説と対立仮説はそれぞれ何か.

H_0: 新薬は有効でない.
H_A: 新薬は有効である.

このようにおくと有効性の検査として自然になる.

信頼区間を用いた仮説検定

それでは上記のように帰無仮説と対立仮説を設定したとき, その仮説検定を実際にどのようにして行うか.

信頼区間を用いた仮説検定

それでは上記のように帰無仮説と対立仮説を設定したとき, その仮説検定を実際にどのようにして行うか.
例えば以下のように信頼区間を使う方法が考えられる:

信頼区間を用いた仮説検定

それでは上記のように帰無仮説と対立仮説を設定したとき, その仮説検定を実際にどのようにして行うか.
例えば以下のように信頼区間を使う方法が考えられる:

集めたデータに基づいて, 帰無仮説中のパラメータ p についての信頼区間を作成し, p がその区間に入らなければ, 帰無仮説 H_0 を棄却し対立仮説 H_A を採用する.

例題: 大卒者への調査結果

初めの問題についての大卒者の回答結果をまとめたデータを作成した. このデータでは, 回答数は 50 人で正答の (c) を選んだ割合は 24%だった. このとき, 信頼区間を作成し回答結果を評価するための仮説を検定せよ.

例題: 大卒者への調査結果

初めの問題についての大卒者の回答結果をまとめたデータを作成した. このデータでは, 回答数は 50 人で正答の (c) を選んだ割合は 24%だった. このとき, 信頼区間を作成し回答結果を評価するための仮説を検定せよ.

ステップ 1 (仮説を立てる):

H_0: $p = 0.33$

H_A: $p \neq 0.33$

例題: 大卒者への調査結果

ステップ 2 (推定値 $\hat{p}$ の分布を求める):
データから比率の推定値は $\hat{p} = 0.24$. 調査対象の独立性は満たされており, また成功・失敗条件も $n\hat{p} = 50 \times 0.24 \geq 10$,
$n(1 - \hat{p}) = 50 \times 0.76 \geq 10$ より満たされている.
よって $\hat{p}$ は正規分布に従う.

例題: 大卒者への調査結果

ステップ 2 (推定値 $\hat{p}$ の分布を求める):
データから比率の推定値は $\hat{p} = 0.24$. 調査対象の独立性は満たされており, また成功・失敗条件も $n\hat{p} = 50 \times 0.24 \geq 10$,
$n(1-\hat{p}) = 50 \times 0.76 \geq 10$ より満たされている.
よって $\hat{p}$ は正規分布に従う.

ステップ 3 (p の信頼区間を求める):
さらに $\hat{p}$ の標準誤差は
$SE = \sqrt{\hat{p}(1-\hat{p})/n} = \sqrt{0.24 \times 0.76/50} = 0.06$. ゆえに正規分布に従うときの 95%信頼区間は
$\hat{p} \pm 1.96 SE$ だったから,

例題: 大卒者への調査結果

正答比率 p の 95%信頼区間は
$0.24 \pm 1.96 \times 0.06 = 0.24 \pm 0.12.$

例題: 大卒者への調査結果

正答比率 p の 95%信頼区間は
$0.24 \pm 1.96 \times 0.06 = 0.24 \pm 0.12.$

ステップ 4 (信頼区間に基づいて仮説検定):
帰無仮説の比率 $p = 0.33$ は 95%信頼区間の範囲に収まっている. よって対立仮説 $p \neq 0.33$ を支持する十分な証拠があるとは言えないので, 帰無仮説は棄却されない. (あてずっぽうと同じである可能性が残る)

意思決定のミス

実は仮説検定はあくまで統計的推測なので，数学的な演繹とは違い，どのようにしても必ずある程度の誤りを含んでしまう．その誤りには以下の二通りがある:

意思決定のミス

実は仮説検定はあくまで統計的推測なので, 数学的な演繹とは違い, どのようにしても必ずある程度の誤りを含んでしまう. その誤りには以下の二通りがある:

(a) 第 1 種の過誤: 帰無仮説が真実なのに帰無仮説を棄却し対立仮説を採用してしまう誤り.

意思決定のミス

実は仮説検定はあくまで統計的推測なので, 数学的な演繹とは違い, どのようにしても必ずある程度の誤りを含んでしまう. その誤りには以下の二通りがある:

(a) 第 1 種の過誤: 帰無仮説が真実なのに帰無仮説を棄却し対立仮説を採用してしまう誤り.

(b) 第 2 種の過誤: 対立仮説が真実なのに帰無仮説を棄却しない誤り.

例題: 裁判

被告への判決を下す裁判では無罪が H_0, 有罪が H_A となる．この時の第 1 種の過誤と第 2 種の過誤をそれぞれ有罪/無罪という言葉を使って表せ．

例題: 感染

あるウイルスへの感染を調べる検査では非感染が H_0, 感染が H_A となる．この時の第 1 種の過誤と第 2 種の過誤はそれぞれ偽陽性と偽陰性のどちらか.

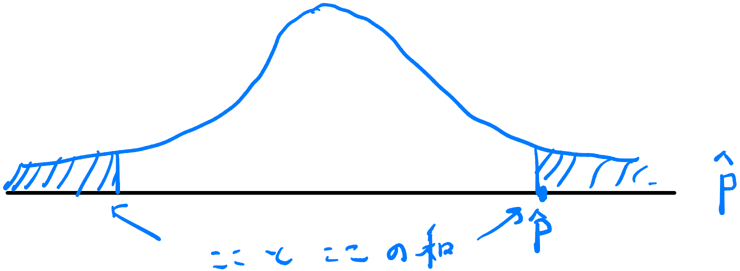
P 値とは

汎用性の高い仮説検定の方法として P 値を使う方法がある.

P 値とは

汎用性の高い仮説検定の方法として P 値を使う方法がある.

P 値とは, 帰無仮説のもとで検定統計量がその値となる確率のことで, 直観的には以下の分布の色付き部分の面積が P 値を表す. (統計量の分布が対称で両側検定の場合):



例題: 石炭の使用

2013 年に行われた 1028 人のアメリカの成人の無作為抽出調査では, 56%が核兵器削減を支持していたことが分かった.

例題: 石炭の使用

2013 年に行われた 1028 人のアメリカの成人の無作為抽出調査では, 56%が核兵器削減を支持していたことが分かった.

このデータを使って, 有意水準 5%で過半数のアメリカ人は核兵器廃絶を支持していたかを検証するための仮説検定を行え.

ステップ 1: 仮説の設定

以下のように帰無仮説と対立仮説を設定する:

ステップ 1: 仮説の設定

以下のように帰無仮説と対立仮説を設定する:

H_0: $p = 0.5$ (支持率に有意差はない)
H_A: $p \neq \cancel{0}$ (支持率に有意差あり)

0.5

ステップ 2: 推定値 $\hat{p}$ の分布

帰無仮説 $p = 0.5$ を仮定する．このとき，
$n = 1028$ であり，成功・失敗条件は

ステップ 2: 推定値 $\hat{p}$ の分布

帰無仮説 $p = 0.5$ を仮定する．このとき，
$n = 1028$ であり，成功・失敗条件は

$$
\begin{aligned}
np &= 1028 \times 0.5 \geq 10 \quad \text{かつ} \\
n(1-p) &= 1028 \times (10.5) \geq 10
\end{aligned}
$$

より満たされる．

ステップ 2: 推定値 $\hat{p}$ の分布

帰無仮説 $p = 0.5$ を仮定する．このとき，
$n = 1028$ であり，成功・失敗条件は

$$
\begin{aligned}
np &= 1028 \times 0.5 \geq 10 \quad \text{かつ} \\
n(1-p) &= 1028 \times (10.5) \geq 10
\end{aligned}
$$

より満たされる.ゆえに帰無仮説のもとで標本比率 $\hat{p}$ は正規分布に従う.

ステップ 2: 推定値 $\hat{p}$ の分布

帰無仮説 $p = 0.5$ を仮定する．このとき，
$n = 1028$ であり，成功・失敗条件は

$$
\begin{aligned}
np &= 1028 \times 0.5 \geq 10 \quad \text{かつ} \\
n(1-p) &= 1028 \times (10.5) \geq 10
\end{aligned}
$$

より満たされる.ゆえに帰無仮説のもとで標本比率 $\hat{p}$ は正規分布に従う．

またこの仮説のもとで，標準誤差は
$SE = \sqrt{p(1-p)/n} = \sqrt{0.5 \times 0.5/1028} = 0.015.$

ステップ 3: 裾野の確率

データから標本比率は $\hat{p} = 0.56$. このときの $\hat{p}$ の値の裾野の確率を求める.

ステップ 3: 裾野の確率

データから標本比率は $\hat{p} = 0.56$. このときの $\hat{p}$ の値の裾野の確率を求める.

Z スコアを計算すると,

$$Z = \frac{\hat{p} - p}{SE} = \frac{0.56 - 0.5}{0.015} = 4.$$

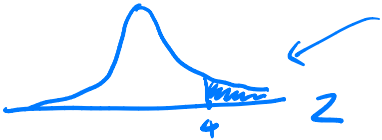
ステップ 3: 裾野の確率

データから標本比率は $\hat{p} = 0.56$. このときの $\hat{p}$ の値の裾野の確率を求める.

Z スコアを計算すると,

$$Z = \frac{\hat{p} - p}{SE} = \frac{0.56 - 0.5}{0.015} = 4.$$

よって右側の裾野の確率は 0.0002 以下.



ステップ 4: P 値を使って仮説検定

今の仮説検定は両側検定であったから, P 値は裾野の確率の二倍で 0.0004 以下となる.

ステップ 4: P 値を使って仮説検定

今の仮説検定は両側検定であったから, P 値は裾野の確率の二倍で 0.0004 以下となる.

よって P 値は有意水準 $\alpha = 0.05$ 以下であるから, 帰無仮説のもとではこのような結果はほぼ出ないということで, 帰無仮説は棄却される. つまり対立仮説, 支持率には有意な差があり過半数が削減を支持している, が採用される.