

『AI・データサイエンスのための数学入門』学習補助資料

仮説検定を、判断の手順として理解する

仮説・確率・結論を区別する補助ノート

1 検定が答える問いを明確にする

仮説検定は、母数について置いた仮説と、得られたデータとの整合性を調べる手続きです。例えば平均重量について、 $H_0: \mu = 500$ を基準に、今回の標本平均がどの程度外れているかを調べます。

背理法との違い

背理法では仮定から論理的矛盾を導きます。検定では、帰無仮説のもとで**確率が小さい結果**を根拠に棄却します。まれな結果も起こるため、棄却が論理的な反証になるわけではありません。

検定の 5 段階

1. 対象と仮説を定める。誰の何を推測するのか、母数と基準値を明記します。
2. 検定方法と有意水準を先に決める。両側か片側かも、結果を見る前に選びます。
3. 前提を確認する。独立性、抽出方法、分布や近似の条件を確認します。
4. 統計量と p 値を計算する。帰無仮説のもとでの分布と比較します。
5. 結論と大きさを報告する。棄却の判断に加え、推定値、信頼区間、効果の大きさを示します。

帰無仮説は「効果なし」に限らない

帰無仮説とは、検定で基準にする母数や分布についての主張です。 $H_0: \mu = 500$ のほか、適合度検定なら「各カテゴリーの確率は指定された値である」となります。「効果なし」は典型例です。

右側検定では、厳密には $H_0: \mu \leq \mu_0$ 、 $H_1: \mu > \mu_0$ と置くこともあります。通常の母平均 Z 検定では、帰無仮説の境界 $\mu = \mu_0$ で棄却確率が最大になるため、そこで確率を計算します。

「偶然だけ」と「効果だけ」の二択ではない

真の差があっても標本には偶然の誤差が加わります。また、標本の選び方や測定に偏りがあれば、差にはそれらの影響も含まれます。検定の式だけで原因は確定できません。

2 p 値は「裾の面積」である

帰無仮説のもとで $Z \sim N(0, 1)$ とし、観測値が $z_{\text{obs}} = 2$ だったとします。 Φ を標準正規分布の分布関数とすると、

	p 値の計算	
$\mu > \mu_0$	$1 - \Phi(2)$	0.0228
$\mu < \mu_0$	$\Phi(2)$	0.9772
$\mu \neq \mu_0$	$2\{1 - \Phi(2)\}$	0.0455

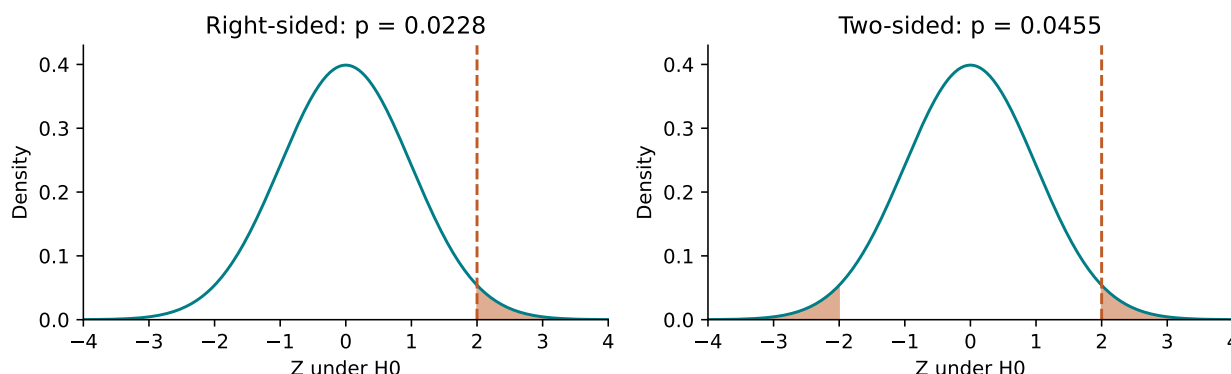


図 1：同じ観測値でも、どちら向きの外れ方を調べるかで p 値が変わります。色の部分は右側検定と両側検定の p 値です。

なぜ観測値「以上」を数えるのか

連続分布では $P(Z = 2) = 0$ です。この一点の確率では、外れ方の程度を比較できません。そこで、検定で決めた方向について、観測値と同程度以上に極端な範囲の確率を求めます。

確率の条件を逆にしない

p 値は、帰無仮説と検定の前提を置いたときのデータの極端さを表します。 $P(H_0 \mid \text{データ})$ でも、「今回の結論が誤りである確率」でもありません。

同じ p 値でも有意水準で判断が変わる

$p = 0.03$ なら、 $\alpha = 0.05$ では棄却し、 $\alpha = 0.01$ では棄却しません。観測後に都合よく基準を変更してはいけません。

両側検定の「絶対値を取って両裾を数える」方法は、このような対称な統計量で使います。非対称な分布や離散分布では、両側の極端さの定義を検定ごとに確認します。

3 過誤と検出力は、繰り返しの確率

判断	H_0 が真	指定した H_1 の状態が真
棄却する	第一種の過誤	検出：確率 $1 - \beta$
棄却しない	正しい判断	第二種の過誤：確率 β

有意水準 α は、 H_0 が真のときに棄却する確率を抑える基準です。正確な連続分布の検定では等号が成り立つことが多い一方、離散検定では α 以下になることもあります。正規近似を用いる検定では、制御も近似的です。

検出力は、真の差の大きさによって変わる

右側 Z 検定で $\alpha = 0.05$ なら、 $Z \geq 1.645$ で棄却します。真の平均を μ_1 とすると、正規母集団・既知の σ のもとで

$$Z \sim N(\lambda, 1), \quad \lambda = \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}, \quad 1 - \beta = 1 - \Phi(1.645 - \lambda).$$

例えば $\lambda = 2$ なら検出力は約 0.639 です。

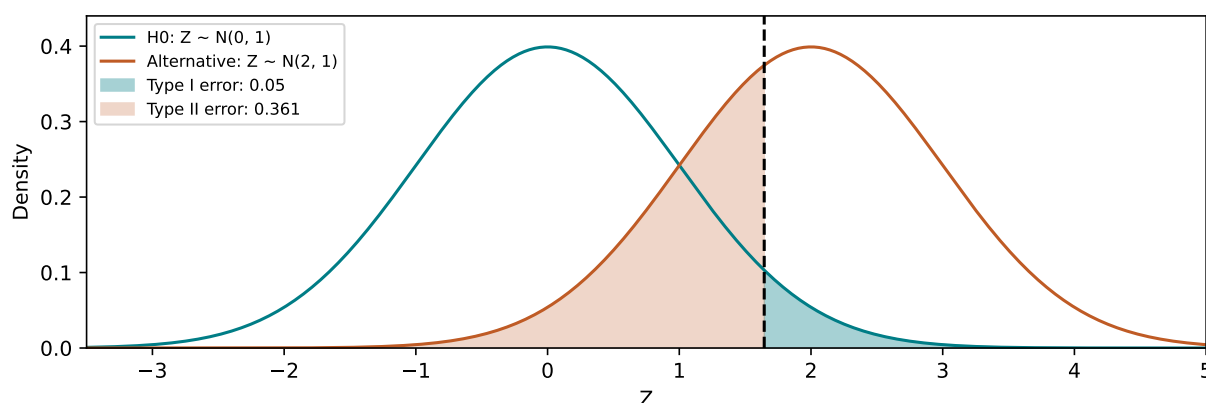


図 2：青は帰無仮説、橙は具体的な対立状態の分布。縦線より右で棄却します。橙の左側の面積が見逃しの確率 β です。

有意差と差の大きさを分ける

$\mu_0 = 100$ 、 $\sigma = 10$ 、観測差 $\bar{x} - \mu_0 = 1$ とします。

$$n = 25 : z = 0.5, p \approx 0.617; \quad n = 2500 : z = 5, p \approx 5.73 \times 10^{-7}.$$

両側検定の有意性は変わりますが、差はどちらも 1、 σ で標準化した差は 0.1 です。大きな標本は小さな差も検出します。2 群の効果量では、分母に使う標準偏差の定義も明記します。

4 母平均・母比率：何で割っているのか

母平均：個々の値の散らばりではなく標本平均の散らばり

独立同分布の標本では、 $E[\bar{X}] = \mu$ 、 $V[\bar{X}] = \sigma^2/n$ です。したがって、

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

は「標本平均が基準値から標準誤差何個分ずれたか」を表します。 σ で割ることと、 $\sigma/\sqrt{n}$ で割ることを区別しましょう。

母分散が未知で正規母集団からの独立標本なら、 $s^2 = \sum (X_i - \bar{X})^2 / (n-1)$ として

$$T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t_{n-1} \quad (H_0 \text{ のもと})$$

を使います。分散を推定する不確実性が加わるため、有限標本では Z 分布と同じではありません。

母比率：分母に p_0 を入れる理由

購入なら $B_i = 1$ 、非購入なら $B_i = 0$ とすると、 $\hat{p} = n^{-1} \sum B_i$ です。 $H_0 : p = p_0$ のもとでは

$$E[B_i] = p_0, \quad V[B_i] = p_0(1-p_0), \quad V[\hat{p}] = \frac{p_0(1-p_0)}{n}.$$

つまり、**帰無仮説が指定するばらつき**で標準化するので、分母には $\hat{p}$ でなく p_0 を用います。成功・失敗の期待個数が小さい場合には二項分布による検定などを検討します。「5 以上」は近似精度の保証ではありません。

信頼区間と検定の対応

同じ正規モデルの両側 5% の Z 検定と 95% 信頼区間では、

$$|z| > 1.96 \iff \mu_0 \notin \left[\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}} \right].$$

端点に厳密に一致するときの扱いは、不等号の約束を合わせます。

手順が対応していることが必要

母比率で p_0 を使う検定と、 $\hat{p}$ を使う単純な正規近似信頼区間は、厳密に同じ手順の逆算ではありません。異なる方法を混ぜると、判定が一致しない場合があります。

5 カイ二乗検定：人数のずれを標準化する

適合度と独立性の違い

	適合度検定	独立性の検定
問い	1 つの分類の割合が指定値に合うか	2 つの分類に関連があるか
期待度数	$E_i = np_{i,0}$	$E_{ij} = n_{i.}n_{.j}/n$
自由度	確率が事前指定なら $k - 1$	$r \times c$ 表なら $(r - 1)(c - 1)$

期待度数が小数でもよい理由

独立なら「行 i かつ列 j 」の確率は両者の確率の積です。周辺比率で推定すると、

$$n \cdot \frac{n_{i.}}{n} \cdot \frac{n_{.j}}{n} = \frac{n_{i.}n_{.j}}{n}.$$

期待度数は繰り返したときの平均的な人数なので、37.5 のような小数でも問題ありません。

なぜ平方して、期待度数で割るのか

$\sum(O_i - E_i)$ は正負が打ち消し合うため、ずれ全体を測れません。平方すれば打ち消しがなくなります。また、度数の変動の規模に合わせるために E_i で割ります。

ただし、多項分布では $V[O_i] = np_i(1 - p_i)$ で、 $E_i = np_i$ と同じではありません。各項を独立な標準正規変数の平方とみなすのではなく、**度数の制約と依存を含めた漸近理論**がカイ二乗分布による近似を与えます。

自由度は、制約の数を数える

適合度検定では、 k 個の度数の合計が n という制約があるため、自由に動く成分は $k - 1$ 個です。分布のパラメータを同じデータから推定した場合には、通常さらに調整が必要です。

独立性検定では、周辺合計を固定すると、左上の $(r - 1)(c - 1)$ 個を決めれば、残りは行・列の合計から決まります。これが自由度を理解する手がかりです。

右側の裾を使う

$\chi^2 = \sum(O - E)^2/E$ は非負です。どちら向きのずれも平方で大きな値になるため、 $p = P(\chi_\nu^2 \geq \chi_{\text{obs}}^2)$ を使います。ここで p 値を 2 倍にはしません。

6 本文の練習問題：解答と読み方

問 1：重量の両側 Z 検定

$H_0: \mu = 500$ 、 $H_1: \mu \neq 500$ とします。独立標本と、正規性または正規近似の妥当性を前提として、

$$SE = \frac{8}{\sqrt{36}} = \frac{4}{3}, \quad z = \frac{497 - 500}{4/3} = -2.25.$$

$$p = 2\{1 - \Phi(2.25)\} = 2(1 - 0.9878) = \boxed{0.0244}.$$

$0.0244 < 0.05$ なので棄却します。平均重量が仕様値と異なることを示す証拠があります。推定差は -3 g です。対応する 95% 信頼区間は約 $[494.39, 499.61]$ g で、500 g を含みません。

問 2： p 値の解釈

1. $p = 0.03$ なら棄却し、購入率が基準より高いことを支持します。ただし、広告による因果効果の判断には、比較条件や割付方法の検討も必要です。
2. $p = 0.08$ なら棄却できません。「上昇を示す証拠が 5% 水準では十分でない」と述べ、「効果がない」とは断定しません。
3. 事前に定めた α に対して、 $p \leq \alpha$ なら棄却し、 $p > \alpha$ なら棄却しません。
4. p 値は、帰無仮説と検定の前提のもとで、観測値と同程度以上に対立仮説の方向へ外れた統計量が得られる確率です。

問 3：購入率の右側検定

$$\hat{p} = \frac{30}{200} = 0.15, \quad H_0: p = 0.10, \quad H_1: p > 0.10,$$

$$z = \frac{0.15 - 0.10}{\sqrt{0.10 \cdot 0.90/200}} \approx \boxed{2.357} > 1.645.$$

帰無仮説を棄却します。右側 p 値は約 0.0092 です。帰無仮説での成功・失敗の期待個数は 20 と 180 です。ここでの結論は、基準の 10% を既知と扱った正規近似によります。

7 度数の計算と、結論の伝え方

本文・問 4：期待度数とずれ

行合計はいずれも 100、列合計は 62 と 138、総数は 200 です。

$$E = \begin{pmatrix} 31 & 69 \\ 31 & 69 \end{pmatrix}, \quad O - E = \begin{pmatrix} -7 & 7 \\ 7 & -7 \end{pmatrix}.$$

広告 B の購入者は 38 人で、期待度数 31 人より 7 人多くなっています。

発展：連続性補正なしの Pearson のカイ二乗検定まで進めると、

$$\chi^2 = 2 \left(\frac{49}{31} + \frac{49}{69} \right) \approx 4.582, \quad \nu = 1, \quad p \approx 0.0323.$$

5% 水準で独立性を棄却します。ただし、期待度数より多いことだけでは有意とはいえず、統計量とその分布で判断します。

報告文の例

購入率は広告 A で 24%、広告 B で 38% で、観測された差は 14 パーセントポイントであった。Pearson のカイ二乗検定（連続性補正なし）では、広告の種類と購入の有無に関連が認められた ($\chi^2(1) = 4.582, p = 0.0323$)。この関連だけから広告の因果効果を確定することはできない。

有意でない結果から、同等性は証明できない

「差がある」と示せない理由には、差が小さいことのほかに、データ不足やばらつきの大きさがあります。同等と判断したいなら、実質的に許容する差の範囲を先に定め、同等性を調べる方法が必要です。

検定を多数行うとき

すべての帰無仮説が真で、20 個の独立な検定を各 5% で行くと、少なくとも 1 個を誤って棄却する確率は

$$1 - (1 - 0.05)^{20} \approx \boxed{0.642}.$$

これは各検定の 5% とは違います。例えば Bonferroni の方法では、全体の有意水準を α に抑えるため、 m 個の各検定に α/m を用います。この上界の保証に検定間の独立性は不要です。

8 追加問題と解答

問 1：片側と両側 $z = -1.80$ 、 $\Phi(1.80) = 0.9641$ とする。左側・右側・両側の p 値を求め、5% 水準で判定せよ。

問 2：標本サイズ 母平均 Z 検定で、観測差と母標準偏差を固定したまま n が 4 倍になると、標準誤差と $|z|$ はどうなるか。

問 3：検出力 第 3 頁の右側検定で、真の差が同じまま標本サイズを 4 倍にすると $\lambda = 2$ は 4 になる。このときの検出力を求めよ。 $\Phi(2.355) \approx 0.9907$ としてよい。

問 4：適合度 3 カテゴリーが等確率という仮説に対し、観測度数が $(20, 30, 40)$ だった。期待度数、 χ^2 、自由度を求めよ。自由度 2 の上側 5% 点は 5.991 とする。

問 5：多重検定 10 個の検定に Bonferroni 補正を使い、全体の有意水準を 5% とする。1 個の p 値が 0.008 なら棄却するか。

問 6：読み違い 「 $p = 0.70$ なので、帰無仮説が正しい確率は 70% である」を修正せよ。

解答・解説

1. 左側 0.0359、右側 0.9641、両側 0.0718。事前に左側検定を選んだ場合にのみ 5% で棄却します。結果を見て左側に切り替えてはいけません。
2. 標準誤差は半分、 $|z|$ は 2 倍です。観測差が同じという比較であり、新しい標本の平均が同じになる保証ではありません。
3. $1 - \Phi(1.645 - 4) = \Phi(2.355) \approx 0.9907$ 。真の差が固定されていても、標本サイズにより検出力が変わります。
4. $E = (30, 30, 30)$ 、 $\chi^2 = 100/30 + 0 + 100/30 = 20/3 \approx 6.667$ 、自由度 $3 - 1 = 2$ 。6.667 > 5.991 なので棄却します。
5. 個別基準は $0.05/10 = 0.005$ 。0.008 > 0.005 なので棄却しません。
6. 「帰無仮説のもとで、定めた統計量が観測値と同程度以上に極端となる確率が 70% である」。帰無仮説自体の確率を表す値ではありません。

講義で押さえること

仮説と前提を先に定める。 p 値の条件を逆にしない。非有意を「差がない」と読み替えない。有意性と効果の大きさを分けて報告する。