

統計 (MI)

第2回 t分布(基礎編)

授業担当: 徳永伸一

東京医科歯科大学教養部 数学講座

今日の授業の概要:

- 統計 でやった推定・検定の基礎(第5章)を振り返りつつ
- t 分布を使った推定および検定(t 検定)について

統計 再確認

あらためて「第5章 推定・検定の基礎」について

- 5.1 点推定
 - 推定量の不偏性
- 5.2 母数の区間推定
- 5.3 仮説検定の考え方
- 5.4 片側検定

ポイント:

例題5.1, 5.3(母平均の区間推定と検定)ではいずれも母分散が既知だった.

→もっとも基本的なケースだが, 現実にはあまりありそうにない.

[再確認] 統計 の内容

第6章 推定・検定の実際()

- 6.1 母分散が未知の場合の母平均の推測(t分布)
- 6.2 独立な2標本にもとづく母平均の差に関する推測
- 6.3 等分散の検定(F分布)
- 6.4 対応のある2標本にもとづく母平均の差に関する推測(t分布)
- 6.5 母集団比率に関する推測(正規近似)
- 6.6 適合度検定(χ^2 検定)

ですが...

[再確認] 統計 の内容

以下の順序で説明することにします:

- 6.1 母分散が未知の場合の母平均の推測 (t分布)
- 6.4 対応のある2標本にもとづく母平均の差に関する推測 (t分布)

(今日はここまで)

- 6.2 独立な2標本にもとづく母平均の差に関する推測
- 6.3 等分散の検定 (F分布)
- 6.5 母集団比率に関する推測 (正規近似)
- 6.6 適合度検定 (χ^2 検定)

ではさっそく6.1から

6.1 母分散が未知の場合の母平均の推測 (t分布) その1

「母分散が未知」 「母分散が既知」

再び軽くおさらい:

母分散が既知の場合の母平均 μ の区間推定について (5章の復習)

- $\bar{X} \sim N(\mu, \sigma^2/n)$ と近似できる.
– 注意: 正規母集団を仮定すれば厳密.
- $Z = (\bar{X} - \mu) / (\sigma / \sqrt{n})$ と標準化すると $Z \sim N(0, 1)$
- $P(-z(\alpha/2) \leq Z \leq z(\alpha/2)) = \gamma$
– ただし $\gamma = 1 - \alpha$. $z(\alpha)$ は $P(Z \leq z(\alpha)) = \alpha$ を満たす値.
- $\uparrow$ を同値変形すると

$$P(\bar{X} - z(\alpha/2) \sigma / \sqrt{n} \leq \mu \leq \bar{X} + z(\alpha/2) \sigma / \sqrt{n}) = \gamma$$

よって「 μ の $(100 \times \gamma)\%$ 信頼区間」は

$$(\bar{X} - z(\alpha/2) \sigma / \sqrt{n}, \bar{X} + z(\alpha/2) \sigma / \sqrt{n})$$

6.1 母分散が未知の場合の母平均の推測(t分布) その2

「母分散既知の場合」のポイントをまとめると

- 母分散 σ^2 を用いて標本平均の分布が表せる.
 - 母集団分布が正規分布なら標本平均の分布も正規分布.
 - 正規母集団を仮定せずとも, 標本サイズ n が十分大きければ標本平均の分布は正規分布で近似でき, いずれにしても正規分布の問題に帰着できた.
 - だがその(標本平均が従う)正規分布 $N(\mu, \sigma^2/n)$ は 母分散 σ^2 を用いて表されているのだから, σ^2 の値がわからないと推測できない.
 - 現実には σ^2 は未知のケースが多い!
- 一方...

6.1 母分散が未知の場合の母平均の推測(t分布) その3

- 標本分散 $s^2 := \{ \sum (x_i - \bar{x})^2 \} / (n-1)$
は, 常にわかる(標本データから計算できる).
そこで...
- 方針1: s^2 を σ^2 の近似値として利用.
 - s^2 は σ^2 の(不偏)推定量ですからね.
 - 「近似」を認めてしまえば, 結局またしても正規分布の問題に帰着. 推定の方法は5章とまったく同じ.
(s の値を σ のところに代入するだけ. 同じ公式が使える)
 - 標本サイズが大きければ, 良い(誤差の少ない)近似値であることが期待できる.
 - だが小標本の場合は誤差が無視できないはず.
→ 方針2へ

6.1 母分散が未知の場合の母平均の推測 (t分布) その4

(ここからが本題)

- 方針2: σ^2 を用いずに表せる (s^2 を含む) 統計量を導入する.
以下の定理を利用.

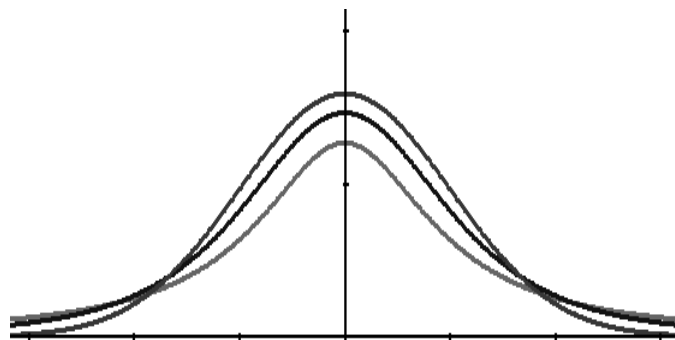
定理6.1

正規母集団 $N(\mu, \sigma^2)$ からの, 大きさ n の無作為標本 $\{X_1, X_2, \dots, X_n\}$ について, 標本平均 $\bar{X}$, 標本分散 (不偏分散) を S^2 とする. このとき, 確率変数

$$T := (\bar{X} - \mu) / (S / \sqrt{n})$$

は自由度 $n-1$ の t 分布 (と呼ばれる分布) に従う.

t分布の密度関数のグラフ



青: 自由度 3, 暗緑: 自由度 1 (赤: 標準正規分布)

6.1 母分散が未知の場合の母平均の推測 (t分布) その5

t分布について

- 「スチューデントのt分布」とも呼ばれる。
 - 「スチューデントstudent」は発案者のペンネーム。
- 厳密に定義するのは少々やっかいなので省略。ともかく理論的に「わかっている」分布として(巻末の数表などを用いて)利用する。
- 密度関数の概形は正規分布と似ている。対称性のある釣鐘型。
- 自由度と呼ばれるパラメータを持つ。
 - サイズ n の標本から求めたスチューデント比 T においては、自由度 ν (ギリシャ文字の「ニュー」) $= n - 1$ 。
 - 自由度の説明は教科書p.72を参照のこと。あまりピンとこなくても、当面は「そういうもの」と割り切って利用すればよいです。
 - ν 大きいほど標準正規分布 $N(0,1)$ に近づき、 $\nu = \infty$ で完全に一致。
 - 自由度 ν のt分布の上側 $100\alpha\%$ 点を $t_{\nu}(\alpha)$ で表す。
 - $t_{\nu}(\alpha)$ の値は教科書p.135の表から。

6.1 母分散が未知の場合の母平均の推測 (t分布) その6

注意と補足

- $T = (X' - \mu) / (S / \sqrt{n})$ はスチューデント比と呼ばれる。
- 「正規母集団」の仮定はt分布を利用する上で本質的に重要。
- 標本平均 X' の標準化変数:
 $Z = (X' - \mu) / (\sigma / \sqrt{n})$ との類似に注意。
 - つまり n が大きければ、 $T \approx Z$ であるということ。
(ただし、今考えているのは n が小さいケースでした)

6.1 母分散が未知の場合の母平均の推測 (t分布) その7

t分布を利用した μ の推定・検定

- 母分散 σ^2 が未知の正規母集団 $N(\mu, \sigma^2)$ からの、大きさ n の無作為標本 $\{x_1, x_2, \dots, x_n\}$ について、標本平均 $\bar{x}$ 、標本分散 (不偏分散) を s^2 とする。

このとき

$$t = (\bar{x} - \mu) / (s / \sqrt{n}) \sim (\text{自由度 } n-1 \text{ の } t \text{ 分布})$$

だから

- 母平均 μ の信頼度 $\gamma = 1 - \alpha$ の信頼区間は

$$(\bar{x} - t_{n-1}(\alpha/2) s / \sqrt{n}, \bar{x} + t_{n-1}(\alpha/2) s / \sqrt{n})$$

- 帰無仮説 $H_0: \mu = \mu_0$ 、対立仮説 $H_1: \mu \neq \mu_0$ に対し、有意水準 α で (両側) 検定するとき、 $\bar{x}$ の棄却域は

$$(\mu_0 - t_{n-1}(\alpha/2) s / \sqrt{n}, \mu_0 + t_{n-1}(\alpha/2) s / \sqrt{n})$$

の補集合。

6.1 母分散が未知の場合の母平均の推測 (t分布) その8

t分布を利用した μ の検定の別解

- もしくは $t_0 = (\bar{x} - \mu) / (s / \sqrt{n})$ とおけば、 t_0 の棄却域は

$$|t_0| \geq t_{n-1}(\alpha/2)$$

と表される。

- 帰無仮説 $H_0: \mu = \mu_0$ 、対立仮説 $H_1: \mu > \mu_0$ に対し、有意水準 α で (片側) 検定するとき、 $\bar{x}$ の棄却域は

$$\bar{x} \geq \mu_0 + t_{n-1}(\alpha) s / \sqrt{n}$$

と表される。また上と同じようにを置けば、 t_0 の棄却域は

$$t_0 \geq t_{n-1}(\alpha)$$

と表される。

[注意] やってることは母分散既知の場合と大差ない。自由度という新たなパラメータが出てきただけ。

6.1 母分散が未知の場合の母平均の推測 (t分布) その9

例題6.1

- 3歳男児の平均身長に関する推定・検定
 - 標本サイズ $n = 30$, 標本平均 $\bar{x}' = 98.4\text{cm}$, 標本分散 (不偏分散) $= 25.2\text{cm}^2$.
 - 教科書巻末の付表Dより $t_{29}(0.025) = 2.045$ を読み取る.
- (1) 「95%信頼区間を求めよ」
- $(\bar{x}' - t_{n-1}(0.025) s / \sqrt{n}, \bar{x}' + t_{n-1}(0.025) s / \sqrt{n})$ に値を代入
- (2) 帰無仮説 $H_0: \mu = 98.6$ を有意水準5%で両側検定.
- 「両側」にする理由: 「片側」にする理由がないから.
 - [解法1] $t_0 = (\bar{x}' - \mu) / (s / \sqrt{n})$ に値を代入し, $t_{29}(0.025) = 2.045$ と比較.
 - [解法2] $\bar{x}'$ の採択域 $(98.6 - t_{n-1}(0.025) s / \sqrt{n}, 98.6 + t_{n-1}(0.025) s / \sqrt{n})$ と $\bar{x}' = 98.4\text{cm}$ を比較.
- (3) 有意水準5%で片側検定.
- 今度は「片側」にする理由 (「発育がよいと言われている」) がある.
 - 教科書巻末の付表Dより $t_{29}(0.05) = 1.699$ を読み取る
 - [解法1] t_0 と比較すればただちに判定可能
 - [解法2] $\bar{x}'$ の棄却域は $\bar{x}' > 98.6 + t_{n-1}(0.05) s / \sqrt{n}$

6.1 母分散が未知の場合の母平均の推測 (t分布) その10

もう一度念押し

- やってることは第5章と大差ない.
- 例題5.1・例題5.3と比較して確認しましょう.

では6.2, 6.3を飛ばして

6.4 対応のある2標本にもとづく母平均の差に関する推測

へ.

6.4 対応のある2標本にもとづく母平均の差に関する推測(1)

- 「対応のある2標本」 「独立な2標本(6.2)」
- 「対応がある」とは(教科書の例):
 - 何組かの一卵性双生児の対を用いてデータを取る.
 - いくつかの資料についての測定結果を対にして考える.
 - 各人のスタート前とゴール直後の体重の差をデータにする.
- 2標本ではあるが, 対データの差をデータとして解析を行うので, 実質的には1標本問題(すなわち6.1で扱ったケースとほぼ同じ解法).
 - 「独立な2標本」の場合(6.2)よりずっと単純!

6.4 対応のある2標本にもとづく母平均の差に関する推測(2)

- 母分散未知のケースを考えている.
 - t分布の利用(正規母集団の仮定が必要).
- n 個の対をなす標本データ $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ から
$$d_1 = x_1 - y_1, d_2 = x_2 - y_2, \dots, d_n = x_n - y_n$$
を算定.
- 2つの母平均をそれぞれ μ_1, μ_2 とし,
$$d' = \sum d_i / n, s^2 = \sum (d_i - d')^2 / (n - 1)$$
とおくと
$$T := (d' - (\mu_1 - \mu_2)) / (s / \sqrt{n})$$
が自由度 $n-1$ の t 分布に従う.
- 以下はこれまでと同じ要領.

6.4 対応のある2標本にもとづく母平均の差に関する推測(3)

- (1) [区間推定] 母平均の差 $\mu_1 - \mu_2$ の信頼係数 $\gamma = 1 - \alpha$ の信頼区間は

$$(d' - t_{n-1}(\alpha/2) s / \sqrt{n}, d' + t_{n-1}(\alpha/2) s / \sqrt{n})$$

- (2) [両側検定]

帰無仮説 $H_0: \mu_1 - \mu_2 = 0$, 対立仮説 $H_1: \mu_1 - \mu_2 \neq 0$
に対し, 有意水準 α で検定する場合,

$$t_0 := d' / (s / \sqrt{n})$$

とおくと, $|t_0| > t_{n-1}(\alpha/2)$ のとき H_0 が有意水準 α で棄却される.

- (3) [片側検定]

帰無仮説 $H_0: \mu_1 - \mu_2 = 0$, 対立仮説 $H_1: \mu_1 - \mu_2 > 0$
に対し, 有意水準 α で検定する場合, $t_0 > t_{n-1}(\alpha)$ のとき H_0 が有意水準 α で棄却される.

6.4 対応のある2標本にもとづく母平均の差に関する推測(4)

例題6.3

無作為に選んだ20人の協力を得て, 1年間食塩の摂取量を控えてもらった。実験を始める前と1年後に最高血圧値を測定し, それぞれに最高血圧値の差を算定したところ, その平均値は4.8mmHg, 不偏分散は12.86mmHg²であった。

- (1) [推定] 減塩による最高血圧値の差の母平均を μ mmHg とするとき, μ の95%信頼区間を求めよ。
(2) [両側検定] 減塩により, 最高血圧値に差が生ずるか否かを, 有意水準5%で検定せよ。
(3) [片側検定] 減塩により, 最高血圧値は下がることはあるが, 上がることはないとするとき, 減塩は血圧を下げる効果があるかどうかを, 有意水準5%で検定せよ。

6.4 対応のある2標本にもとづく母平均の差に関する推測(5)

例題6.3: 解答のポイント

- 「最高血圧値の分布はほぼ正規分布に従うことが知られている(らしい)ので差Dも正規分布に従うと考えられる」
- 母分散が未知なので, (正規母集団を仮定して) t 分布を利用.
- 差の標本平均 $\bar{d} = 4.8$, 不偏分散 $s_d^2 = 12.86$ として t 検定 (1標本の場合と同じ) .