

第 22 回 平均と分散の検定 (12.2, 12.4)

村澤 康友

2024 年 1 月 12 日

今日のポイント

1. 正規母集団の母平均・母分散, 正規母集団の 2 標本問題の母平均の差・母分散の比, ベルヌーイ母集団の母比率の検定の棄却域を求める.
2. H_0 の下で t 分布・ χ^2 分布・F 分布にしたがう検定統計量を, それぞれ t 統計量・ χ^2 統計量・F 統計量という. それらを用いる検定を t 検定・ χ^2 検定・F 検定という.
3. H_0 の下で検定統計量が実現値以上になる確率を p 値 (有意確率) という. p 値 \leq 有意水準なら H_0 を棄却する.

目次

1	正規母集団	1
1.1	母平均の検定 (p. 240)	1
1.2	母分散の検定 (p. 242)	2
1.3	母平均の差の検定 (p. 242)	2
1.4	母分散の比の検定 (p. 244)	3
2	ベルヌーイ母集団 (p. 250)	4
3	p 値	5
4	今日のキーワード	5
5	次回までの準備	5

1 正規母集団

1.1 母平均の検定 (p. 240)

1.1.1 母分散が既知

母集団分布を $N(\mu, \sigma^2)$ とする. ただし σ^2 は既知とする. 次の片側検定問題を考える.

$$H_0 : \mu = c \quad \text{vs} \quad H_1 : \mu > c$$

有意水準を 5 % とする. 大きさ n の無作為標本の標本平均を \bar{X} とすると

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

標準化すると

$$\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sim N(0, 1)$$

検定統計量は

$$Z := \frac{\bar{X} - c}{\sqrt{\sigma^2/n}}$$

H_0 の下で

$$Z \sim N(0, 1)$$

標準正規分布表より H_0 の下で

$$\Pr[Z \geq 1.65] = .05$$

したがって棄却域は $[1.65, \infty)$.

注 1. 棄却域は H_1 に依存する.

- $H_1 : \mu > c$ なら棄却域は $[1.65, \infty)$
- $H_1 : \mu < c$ なら棄却域は $(-\infty, -1.65]$
- $H_1 : \mu \neq c$ なら棄却域は $(-\infty, -1.96] \cup [1.96, \infty)$

例えば $Z = 1.8$ なら $H_0: \mu = c$ は $H_1: \mu > c$ に対しては棄却されるが $H_1: \mu \neq c$ に対しては棄却されない。図 1 を参照。

1.1.2 母分散が未知

標本分散を s^2 とする。 σ^2 を s^2 に置き換えると

$$\frac{\bar{X} - \mu}{\sqrt{s^2/n}} \sim t(n-1)$$

検定統計量は

$$t := \frac{\bar{X} - c}{\sqrt{s^2/n}}$$

H_0 の下で

$$t \sim t(n-1)$$

t 分布表より H_0 の下で、例えば $n = 10$ なら

$$\Pr[t \geq 1.833] = .05$$

したがって棄却域は $[1.833, \infty)$ 。

定義 1. H_0 の下で t 分布にしたがう検定統計量を t 統計量という。

定義 2. t 統計量を用いる検定を t 検定という。

1.2 母分散の検定 (p. 242)

1.2.1 母平均が既知

母集団分布を $N(\mu, \sigma^2)$ とする。ただし μ は既知とする。次の片側検定問題を考える。

$$H_0: \sigma^2 = c \quad \text{vs} \quad H_1: \sigma^2 > c$$

有意水準を 5 % とする。無作為標本 (X_1, \dots, X_n) の標本分散は

$$\hat{\sigma}^2 := \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$$

このとき

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n)$$

検定統計量は

$$\chi^2 := \frac{n\hat{\sigma}^2}{c}$$

H_0 の下で

$$\chi^2 \sim \chi^2(n)$$

χ^2 分布表より H_0 の下で、例えば $n = 10$ なら

$$\Pr[\chi^2 \geq 18.3070] = .05$$

したがって棄却域は $[18.3070, \infty)$ 。

定義 3. H_0 の下で χ^2 分布にしたがう検定統計量を χ^2 統計量という。

定義 4. χ^2 統計量を用いる検定を χ^2 検定という。

1.2.2 母平均が未知

μ が未知なら標本分散は

$$s^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

このとき

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

検定統計量は

$$\chi^2 := \frac{(n-1)s^2}{c}$$

H_0 の下で

$$\chi^2 \sim \chi^2(n-1)$$

χ^2 分布表より H_0 の下で、例えば $n = 10$ なら

$$\Pr[\chi^2 \geq 16.9190] = .05$$

したがって棄却域は $[16.9190, \infty)$ 。

1.3 母平均の差の検定 (p. 242)

1.3.1 母分散が既知

母集団分布を $N(\mu_X, \sigma_X^2), N(\mu_Y, \sigma_Y^2)$ とする。ただし σ_X^2, σ_Y^2 は既知とする。次の片側検定問題を考える。

$$H_0: \mu_X = \mu_Y \quad \text{vs} \quad H_1: \mu_X > \mu_Y$$

有意水準を 5 % とする。各母集団から独立に抽出した大きさ m, n の無作為標本の標本平均を \bar{X}, \bar{Y} とすると

$$\bar{X} \sim N\left(\mu_X, \frac{\sigma_X^2}{m}\right)$$

$$\bar{Y} \sim N\left(\mu_Y, \frac{\sigma_Y^2}{n}\right)$$

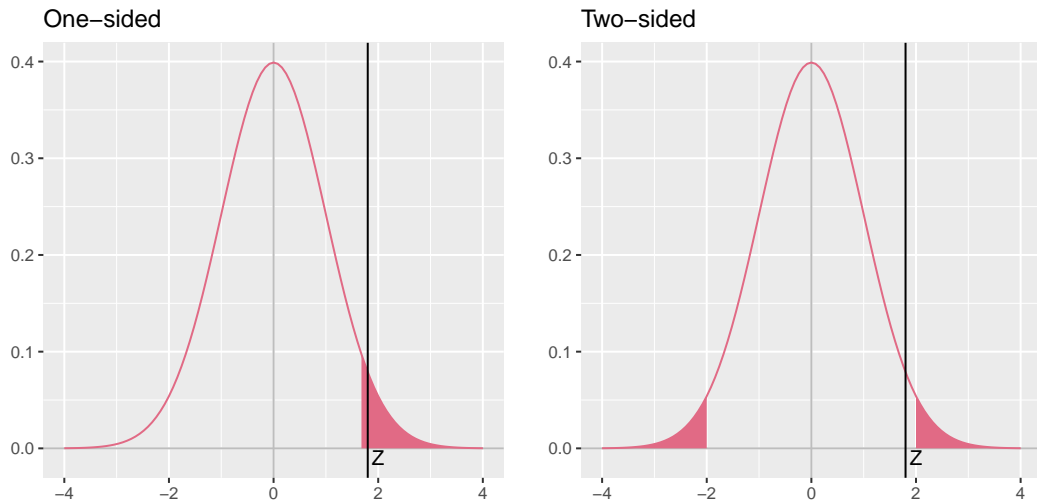


図1 片側検定と両側検定の棄却域

両者は独立だから

$$\bar{X} - \bar{Y} \sim N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}\right)$$

標準化すると

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\sigma_X^2/m + \sigma_Y^2/n}} \sim N(0, 1)$$

検定統計量は

$$Z := \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_X^2/m + \sigma_Y^2/n}}$$

H_0 の下で

$$Z \sim N(0, 1)$$

標準正規分布表より H_0 の下で

$$\Pr[Z \geq 1.65] = .05$$

したがって棄却域は $[1.65, \infty)$.

1.3.2 母分散が未知

$\sigma_X^2 = \sigma_Y^2 = \sigma^2$ なら

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\sigma^2(1/m + 1/n)}} \sim N(0, 1)$$

プールした標本分散を s^2 とする. σ^2 を s^2 に置き換えると

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{s^2(1/m + 1/n)}} \sim t(m + n - 2)$$

検定統計量は

$$t := \frac{\bar{X} - \bar{Y}}{\sqrt{s^2(1/m + 1/n)}}$$

H_0 の下で

$$t \sim t(m + n - 2)$$

t 分布表より H_0 の下で, 例えば $m = 4, n = 6$ なら

$$\Pr[t \geq 1.860] = .05$$

したがって棄却域は $[1.860, \infty)$.

注 2. $\sigma_X^2 \neq \sigma_Y^2$ なら近似的な検定を用いる.

1.4 母分散の比の検定 (p. 244)

1.4.1 母平均が既知

母集団分布を $N(\mu_X, \sigma_X^2), N(\mu_Y, \sigma_Y^2)$ とする. ただし μ_X, μ_Y は既知とする. 次の片側検定問題を考える.

$$H_0 : \sigma_X^2 = \sigma_Y^2 \quad \text{vs} \quad H_1 : \sigma_X^2 > \sigma_Y^2$$

有意水準を 5% とする. 各母集団から独立に抽出した大きさ m, n の無作為標本の標本分散を $\hat{\sigma}_X^2, \hat{\sigma}_Y^2$ とすると

$$\frac{m\hat{\sigma}_X^2}{\sigma_X^2} \sim \chi^2(m)$$

$$\frac{n\hat{\sigma}_Y^2}{\sigma_Y^2} \sim \chi^2(n)$$

両者は独立だから

$$\frac{\hat{\sigma}_X^2/\sigma_X^2}{\hat{\sigma}_Y^2/\sigma_Y^2} \sim F(m, n)$$

すなわち

$$\frac{\hat{\sigma}_X^2/\hat{\sigma}_Y^2}{\sigma_X^2/\sigma_Y^2} \sim F(m, n)$$

検定統計量は

$$F := \frac{\hat{\sigma}_X^2}{\hat{\sigma}_Y^2}$$

H_0 の下で

$$F \sim F(m, n)$$

F 分布表より H_0 の下で, 例えば $m = 4, n = 6$ なら

$$\Pr[F \geq 4.534] = .05$$

したがって棄却域は $[4.534, \infty)$.

定義 5. H_0 の下で F 分布にしたがう検定統計量を **F 統計量** という.

定義 6. F 統計量を用いる検定を **F 検定** という.

1.4.2 母平均が未知

標本分散を s_X^2, s_Y^2 とすると

$$\frac{(m-1)s_X^2}{\sigma_X^2} \sim \chi^2(m-1)$$

$$\frac{(n-1)s_Y^2}{\sigma_Y^2} \sim \chi^2(n-1)$$

両者は独立だから

$$\frac{s_X^2/\sigma_X^2}{s_Y^2/\sigma_Y^2} \sim F(m-1, n-1)$$

すなわち

$$\frac{s_X^2/s_Y^2}{\sigma_X^2/\sigma_Y^2} \sim F(m-1, n-1)$$

検定統計量は

$$F := \frac{s_X^2}{s_Y^2}$$

H_0 の下で

$$F \sim F(m-1, n-1)$$

F 分布表より H_0 の下で, 例えば $m = 4, n = 6$ なら

$$\Pr[F \geq 5.409] = .05$$

したがって棄却域は $[5.409, \infty)$.

2 ベルヌーイ母集団 (p. 250)

母集団分布を $\text{Bin}(1, p)$ とする. 次の片側検定問題を考える.

$$H_0 : p = p_0 \quad \text{vs} \quad H_1 : p > p_0$$

有意水準を 5 % とする. $\text{Bin}(1, p)$ の平均は p , 分散は $p(1-p)$. 大きさ n の無作為標本の標本平均 (= 標本比率) を \hat{p} とすると, 中心極限定理より

$$\hat{p} \stackrel{a}{\sim} N\left(p, \frac{p(1-p)}{n}\right)$$

標準化すると

$$\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \stackrel{a}{\sim} N(0, 1)$$

検定統計量は

$$Z := \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

H_0 の下で

$$Z \stackrel{a}{\sim} N(0, 1)$$

標準正規分布表より H_0 の下で

$$\Pr[Z \geq 1.65] \approx .05$$

したがって近似的な棄却域は $[1.65, \infty)$.

例 1. ある番組の視聴率について有意水準 5 % で次の片側検定を行う.

$$H_0 : p \leq .1 \quad \text{vs} \quad H_1 : p > .1$$

100 世帯を対象に視聴率を調査したら 13 % の視聴率であった. 検定統計量の値は

$$\begin{aligned} Z &:= \frac{.13 - .1}{\sqrt{.1(1-.1)/100}} \\ &= \frac{.03}{\sqrt{.09/100}} \\ &= 1 < 1.65 \end{aligned}$$

したがって H_0 は棄却されない.

3 p 値

定義 7. H_0 の下で検定統計量が実現値以上になる確率を **p 値 (有意確率)** という.

注 3. p 値が有意水準以下なら H_0 を棄却する.

例 2. 有意水準 α の検定を考える. 検定統計量を T , 棄却域を $[t_\alpha, \infty)$, T の実現値を t とすると,

$$\begin{aligned} t \geq t_\alpha &\iff \Pr[T \geq t | H_0] \leq \Pr[T \geq t_\alpha | H_0] \\ &\iff p \leq \alpha \end{aligned}$$

したがって $p \leq \alpha$ なら H_0 は棄却 (図 2).

4 今日のキーワード

母平均の検定 (母分散が既知・未知), t 統計量, t 検定, 母分散の検定 (母平均が既知・未知), χ^2 統計量, χ^2 検定, 母平均の差の検定 (母分散が既知・未知), 母分散の比の検定 (母平均が既知・未知), F 統計量, F 検定, 母比率の検定, p 値 (有意確率)

5 次回までの準備

提出 宿題 6

復習 教科書第 12 章 2 節, 復習テスト 22

予習 教科書第 12 章 3–4 節

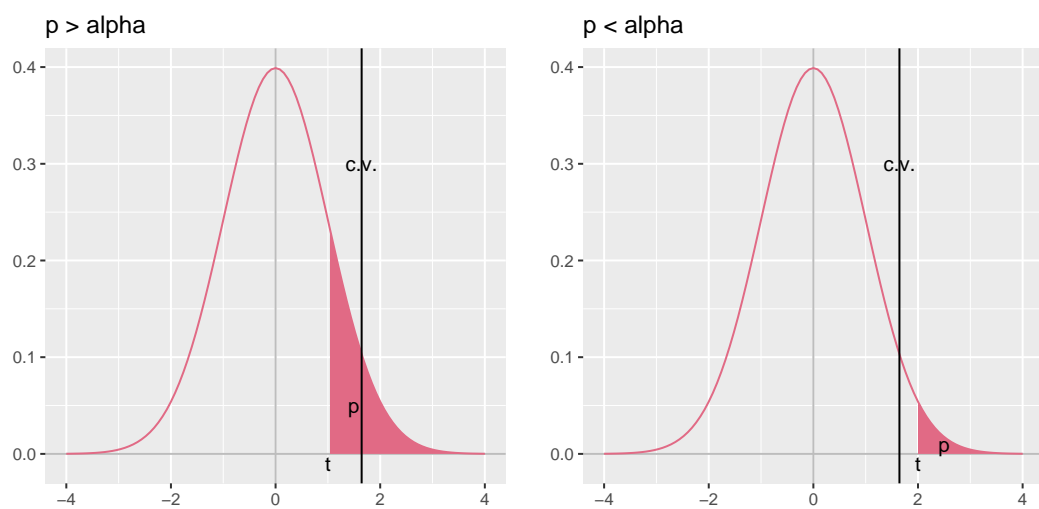


図2 検定統計量と p 値