Lecture 10

母集団の比較

母平均の差の検定 ― 2つの母集団の母平均に差があるかを検定する

※ 正規母集団に限る（一般の母集団はやや高度）

母集団 1

$N(\mu_1, \sigma_1^2)$

無作為復元抽出

標 本 1

大きさ n_1

標本平均 $\bar{x}$

母集団 2

$N(\mu_2, \sigma_2^2)$

無作為復元抽出

標 本 2

大きさ n_2

標本平均 $\bar{y}$

【課題】

得られた標本平均 $\bar{x}, \bar{y}$ をもとに

母平均が等しいかどうかを検定

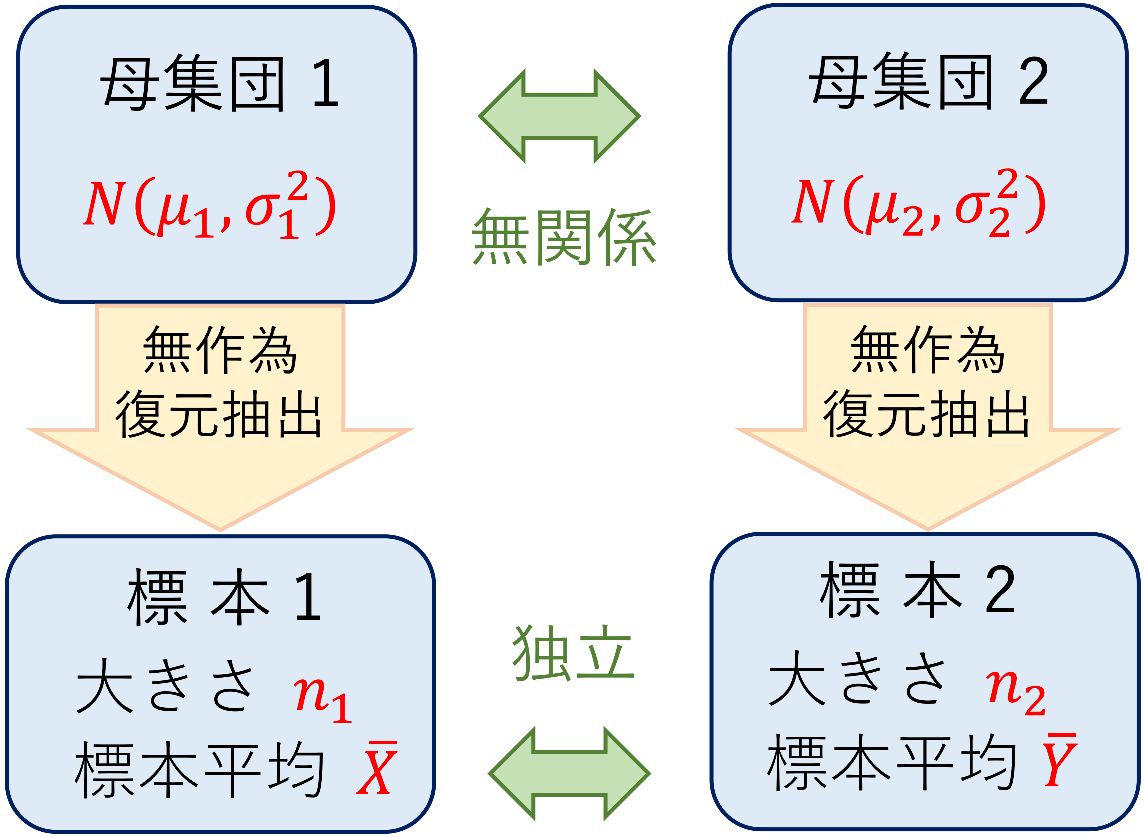
1) 母分散既知の場合

2) 二項母集団の母比率

3) 等分散の場合

母平均の差の検定（母分散既知）

※ 準備



標本平均を確率変数 $\bar{X}, \bar{Y}$ として扱い，$\bar{X} - \bar{Y}$ の分布を調べる

$$\bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$$

$$\bar{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

$\bar{X} - \bar{Y}$ は？

（復習）確率変数の和

- 一般の確率変数 $X,\ Y$ と定数 a, b に対して

（1）［平均値の線形性］$E[aX + bY] = aE[X] + bE[Y]$

（2）分散は線形性をもたないが $V[aX] = a^2V[X]$

- 確率変数 $X,\ Y$ が独立ならば

（3）［平均値の乗法性］$E[XY] = E[X]E[Y]$

（4）［分散の加法性］$V[X + Y] = V[X] + V[Y]$

- 確率変数 $X \sim N(\mu_1, \sigma_1^2),\ Y \sim N(\mu_2, \sigma_2^2)$ が独立ならば

（5）線形結合 $aX + bY$ も正規分布に従い，

$$aX + bY \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$$

定理　（標本平均の差の分布）

正規母集団 $N(\mu_1, \sigma_1^2)$ から取り出した n_1 個の無作為標本の標本平均を $\bar{X}_1$，別の正規母集団 $N(\mu_2, \sigma_2^2)$ から取り出した n_2 個の無作為標本の標本平均を $\bar{X}_2$ とするとき，

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2 \,, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

母集団 1
$N(\mu_1, \sigma_1^2)$

標本 1
大きさ n_1
標本平均 $\bar{X}_1$

$\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$

無関係

独立

母集団 2
$N(\mu_2, \sigma_2^2)$

標本 2
大きさ n_2
標本平均 $\bar{X}_2$

$\bar{X}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$

$\bar{X}_1 - \bar{X}_2$ について

平均値

$\mu_1 - \mu_2$

分散

$\frac{\sigma_1^2}{n_1} + (-1)^2 \frac{\sigma_2^2}{n_2}$

母平均の差の検定（母分散既知）

母集団 1
$N(\mu_1, \sigma_1^2)$

無関係

母集団 2
$N(\mu_2, \sigma_2^2)$

標本 1
大きさ n_1
標本平均 $\bar{X}_1$

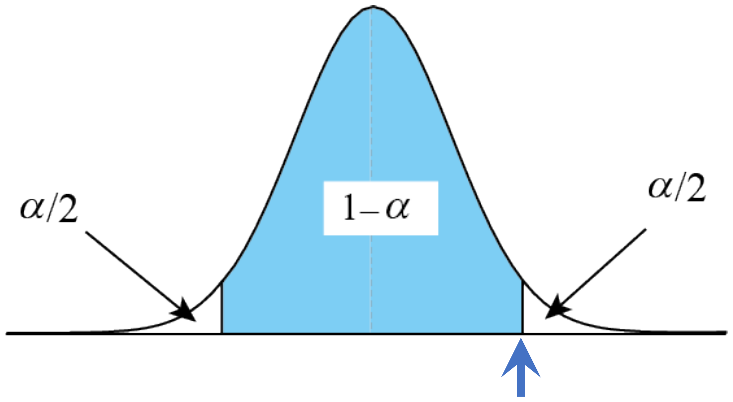
独立

標本 2
大きさ n_2
標本平均 $\bar{X}_2$

$$\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right) \qquad \bar{X}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2\ , \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$



上側 $\boldsymbol{\alpha/2}$点 $= \boldsymbol{z(\alpha/2)} =$ 両側 $\boldsymbol{\alpha}$点

例題 10.1

2つの地区 A, B に生育するある植物の背丈は正規分布に従っており, A地区, B地区それぞれの標準偏差は 4.2 と 3.8 である. A地区からサンプリングした24個体の背丈の平均値は85.8, B地区からサンプリングした35個体の背丈の平均値は 88.2 であった. A地区, B地区で背丈の違いは認められるだろうか.

例題 10.1　2つの地区 A, B に生育するある植物の背丈は正規分布に従っており, A地区, B地区それぞれの標準偏差は 4.2 と 3.8 である. A地区からサンプリングした24個体の背丈の平均値は85.8, B地区からサンプリングした35個体の背丈の平均値は 88.2 であった. A地区, B地区で背丈の違いは認められるだろうか.

A地区, B地区の植物の平均値をそれぞれ μ_1, μ_2 とおく.

母集団 A
$N(\mu_1, \sigma_1^2)$
$\sigma_1 = 4.2$

母集団 B
$N(\mu_2, \sigma_2^2)$
$\sigma_2 = 3.8$

母分散既知

標本 A
大きさ $n_1 = 24$
標本平均 $\bar{X}_1$

標本 B
大きさ $n_2 = 35$
標本平均 $\bar{X}_2$

$$\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right), \qquad \bar{X}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

例題 10.1　2つの地区 A, B に生育するある植物の背丈は正規分布に従っており, A地区, B地区それぞれの標準偏差は 4.2 と 3.8 である. A地区からサンプリングした24個体の背丈の平均値は85.8, B地区からサンプリングした35個体の背丈の平均値は 88.2 であった. A地区, B地区で背丈の違いは認められるだろうか.

A地区, B地区の植物の平均値をそれぞれ μ_1, μ_2 とおく.

母集団 A
$N(\mu_1, \sigma_1^2)$
$\sigma_1 = 4.2$

母集団 B
$N(\mu_2, \sigma_2^2)$
$\sigma_2 = 3.8$

母分散既知

標本 A
大きさ $n_1 = 24$
標本平均 $\bar{X}_1$

標本 B
大きさ $n_2 = 35$
標本平均 $\bar{X}_2$

帰無仮説と対立仮説

$$H_0\colon \mu_1 = \mu_2 \qquad H_1\colon \mu_1 \neq \mu_2$$

検定統計量

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{4.2^2}{24} + \dfrac{3.8^2}{35}}}$$

$$= \frac{\bar{X}_1 - \bar{X}_2}{1.071} \sim N(0,1)$$

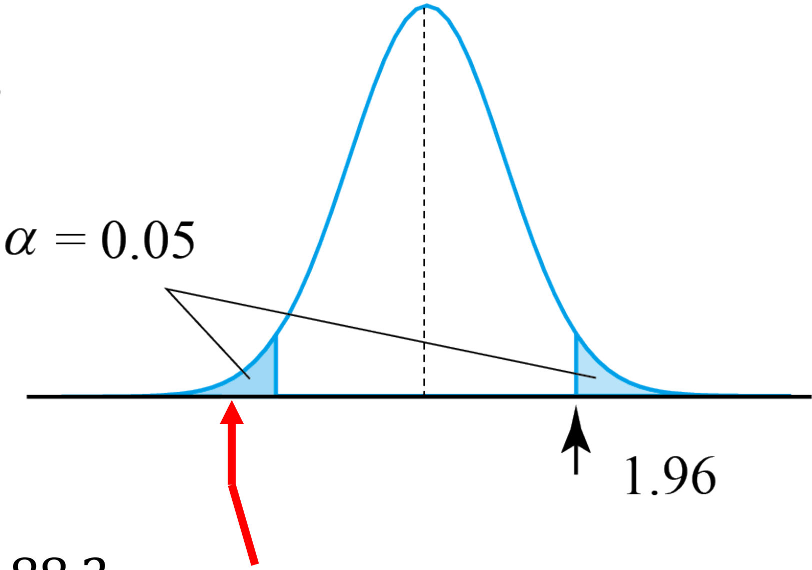
例題 10.1　2つの地区 A, B に生育するある植物の背丈は正規分布に従っており, A地区, B地区それぞれの標準偏差は 4.2 と 3.8 である. A地区からサンプリングした24個体の背丈の平均値は85.8, B地区からサンプリングした35個体の背丈の平均値は 88.2 であった. A地区, B地区で背丈の違いは認められるだろうか.

A地区, B地区の植物の平均値をそれぞれ μ_1, μ_2 とおく.

帰無仮説と対立仮説

$$H_0\colon \mu_1 = \mu_2 \qquad H_1\colon \mu_1 \neq \mu_2$$

有意水準　$\alpha = 0.05$



検定統計量

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{4.2^2}{24} + \dfrac{3.8^2}{35}}}$$

$$= \frac{\bar{X}_1 - \bar{X}_2}{1.071} \sim N(0,1)$$

実現値　$z = \dfrac{85.8 - 88.2}{1.071} = -2.24$

結論　有意水準 $\alpha = 0.05$ の両側検定によって H_0 は棄却される.

母比率の差の検定（二項母集団）

母集団 1
母比率 p_1

無関係

母集団 2
母比率 p_2

標本 1
大きさ n_1
標本比率 $\hat{p}_1$

独立

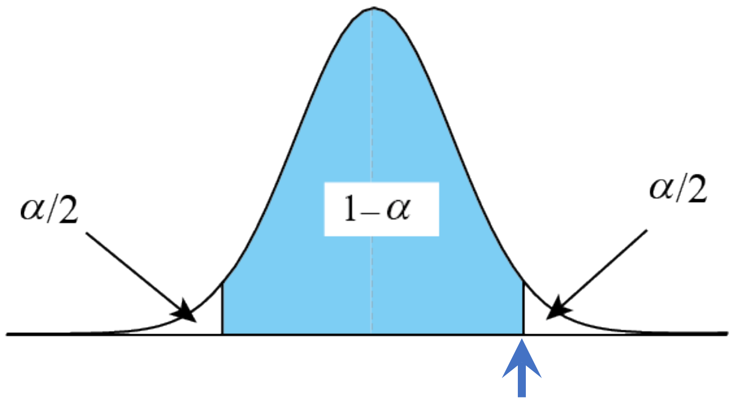
標本 2
大きさ n_2
標本比率 $\hat{p}_2$

$$\hat{p}_1 \sim N\left(p_1\,, \frac{p_1(1-p_1)}{n_1}\right)$$

$$\hat{p}_2 \sim N\left(p_2\,, \frac{p_2(1-p_2)}{n_2}\right)$$

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)$$

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



上側 $\boldsymbol{\alpha/2}$点 $= \boldsymbol{z(\alpha/2)} =$ 両側 $\boldsymbol{\alpha}$点

例題 10.2 （母比率の差）

ある意見項目に対する賛否を男女別に集計したところ，次の結果を得た．賛成者の比率に男女差があるといえるか．

	賛成	反対	計
男	58 (0.592)	40 (0.408)	98 (1.000)
女	28 (0.394)	43 (0.606)	71 (1.000)

男, 女の賛成の母比率をそれぞれ p_1, p_2 とおく.

男 母比率 p_1 → 標本 1 大きさ n_1 標本比率 $\hat{p}_1$

女 母比率 p_2 → 標本 2 大きさ n_2 標本比率 $\hat{p}_2$

無関係

$$\hat{p}_1 \sim N\left(p_1, \frac{p_1(1-p_1)}{n_1}\right)$$

独立

$$\hat{p}_2 \sim N\left(p_2, \frac{p_2(1-p_2)}{n_2}\right)$$

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)$$

稲垣他の本(pp.138-140)の記述は不適切

帰無仮説と対立仮説　$H_0\colon p_1 = p_2 = p$　$H_1\colon p_1 \neq p_2$

有意水準　$\alpha = 0.05$

検定統計量　$\hat{p}_1 - \hat{p}_2 \sim N\left(0, p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right) = N(0, 0.078^2)$

p を実現値から推定　$n_1 = 98$　$n_2 = 71$　$p = \frac{58+28}{98+71} = 0.509$

	賛成	反対	計
男	58 (0.592)	40 (0.408)	98 (1.000)
女	28 (0.394)	43 (0.606)	71 (1.000)

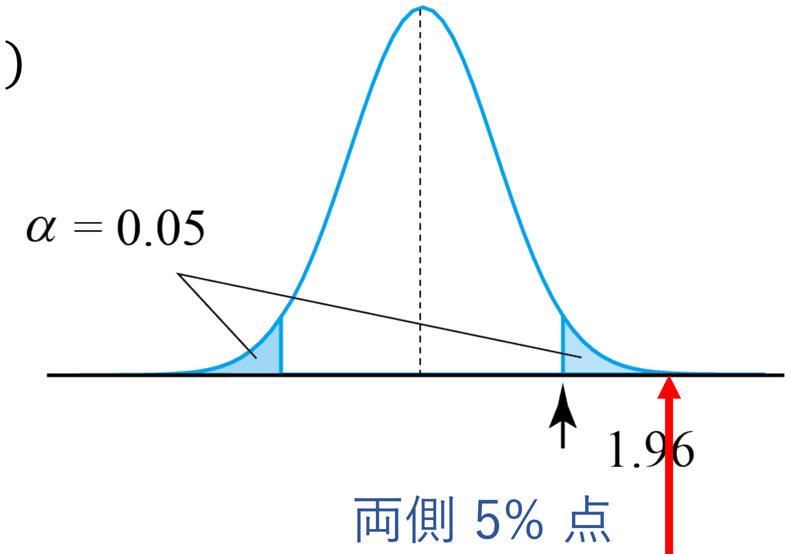
帰無仮説と対立仮説　$H_0\colon p_1 = p_2 = p$　　$H_1\colon p_1 \neq p_2$

有意水準　$\alpha = 0.05$

検定統計量

$$\hat{p}_1 - \hat{p}_2 \sim N\left(0, p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right) = N(0, 0.078^2)$$

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{0.078} \sim N(0,1)$$



実現値　$\hat{p}_1 = \frac{58}{98} = 0.592$　　$\hat{p}_2 = \frac{28}{71} = 0.394$

$$z = \frac{0.592 - 0.394}{0.078} = 2.54$$

結論　有意水準 $\alpha = 0.05$ の両側検定によって H_0 は棄却される．

例題 10.3

インフルエンザについて, 予防接種の有無と感染の有無のデータが次のようであった. 予防接種の効果はあるといえるか. 有意水準 1% で検定せよ.

	感染	非感染
接種	18	67
非接種	45	65

接種, 非接種群の感染率（母比率）をそれぞれ p_1, p_2 とおく.

帰無仮説と対立仮説　$H_0\colon p_1 = p_2 = p$　$H_1\colon p_1 \neq p_2$

有意水準　$\alpha = 0.01$

検定統計量　$\hat{p}_1 - \hat{p}_2 \sim N\left(0, p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$

	感染	非感染	合計
接種	18	67	85
非接種	45	65	110
合計	63	132	195

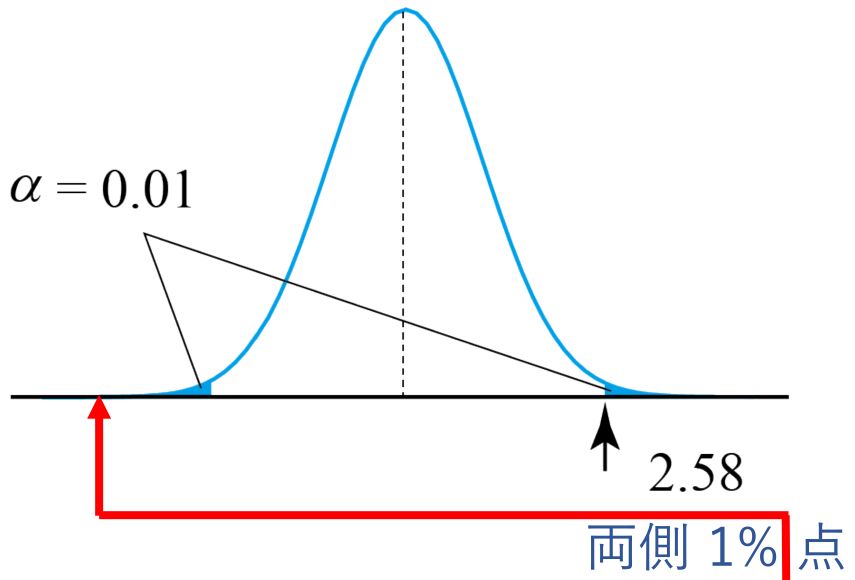
$n_1 = 85$

$n_2 = 110$

$$p = \frac{18 + 45}{85 + 110} = 0.323$$

$$\hat{p}_1 - \hat{p}_2 \sim N(0, 0.0675^2)$$

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{0.067} \sim N(0,1)$$



実現値

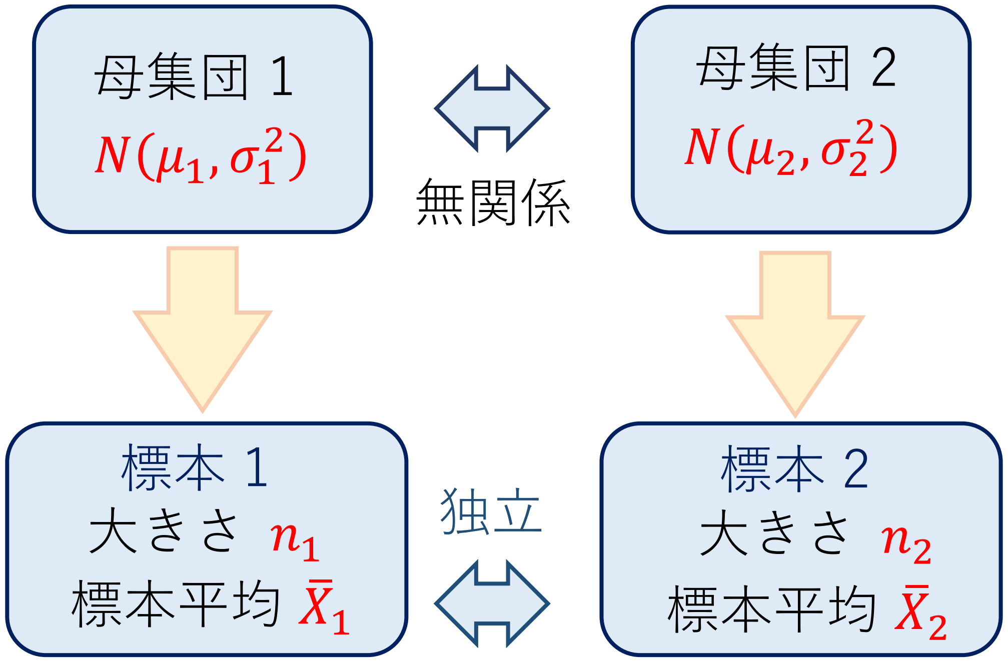
$$\hat{p}_1 = \frac{18}{85} = 0.212$$

$$\hat{p}_2 = \frac{45}{110} = 0.409$$

$$z = \frac{0.212 - 0.409}{0.0675} = -2.91$$

結論　有意水準 $\alpha = 0.01$ の両側検定によって H_0 は棄却される.（高度に有意である.）

母平均の差の検定（母分散未知だが等分散）



$$\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right) \qquad \bar{X}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2\ , \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

- この式は一般に正しいが, 母分散未知の場合は先に進めない.
- しかし、母分散が等しい（等分散）ときは t 検定に持ち込める

定理（等分散のとき, 標本平均の差）

正規母集団 $N(\mu_1, \sigma^2)$ から取り出した n_1 個の無作為標本の標本平均を $\bar{X}_1$, 不偏分散を U_1^2, 別の正規母集団 $N(\mu_2, \sigma^2)$ から取り出した n_2 個の無作為標本の標本平均を $\bar{X}_2$, 不偏分散を U_2^2 とするとき,

$$T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\frac{(n_1 - 1)U_1^2 + (n_2 - 1)U_2^2}{n_1 + n_2 - 2}}} \sim t_{n_1+n_2-2}$$

※ 証明は数理統計学の教科書を参照

例題 10.4（等分散の場合）

ある作物の栽培方式 A, B を比較するため, 畑を15分割して 9 か所で方式 A を, 残りの 6 か所では方式 B を用いて栽培実験を行った. その結果, それぞれの方式について1か所当たりの収穫量(kg)の平均値と不偏分散は

$$\bar{x}_A = 7.32, \quad u_A^2 = 0.44^2, \quad \bar{x}_B = 7.86, \quad u_B^2 = 0.37^2$$

であった. 両者の収穫量は分散の等しい正規分布 $N(\mu_A, \sigma^2)$ と $N(\mu_B, \sigma^2)$ に従うとみなしてよい. 両者の収穫高に有意差は見られるか.

母集団 A
$N(\mu_A, \sigma^2)$

標本 A
大きさ $n_A = 9$
標本平均 $\bar{X}_A$

母集団 B
$N(\mu_B, \sigma^2)$

標本 B
大きさ $n_A = 6$
標本平均 $\bar{X}_B$

帰無仮説と対立仮説

$$H_0 \colon \mu_A = \mu_B \qquad H_1 \colon \mu_A \neq \mu_B$$

有意水準　$\alpha = 0.01$

検定統計量

$$T = \frac{\bar{X}_A - \bar{X}_B - (\mu_A - \mu_B)}{\sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B}\right)\frac{(n_A - 1)U_A^2 + (n_B - 1)U_B^2}{n_A + n_B - 2}}}$$

$$= \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\left(\frac{1}{9} + \frac{1}{6}\right)\frac{8U_A^2 + 5U_B^2}{13}}} \sim t_{13}$$

実現値

$$t = \frac{7.32 - 7.86}{\sqrt{\left(\frac{1}{9} + \frac{1}{6}\right)\frac{8 \cdot 0.44^2 + 5 \cdot 0.37^2}{13}}} = -2.472$$

$|t| \geq t_{13}(0.025) = 2.160$　上側 2.5% 点

結論　有意水準 $\alpha = 0.05$ の両側検定によって H_0 は棄却される．

母平均の差の検定（母分散未知）

母集団 1
$N(\mu_1, \sigma_1^2)$

標本 1
大きさ n_1
標本平均 $\bar{X}_1$

無関係

母集団 2
$N(\mu_2, \sigma_2^2)$

標本 2
大きさ n_2
標本平均 $\bar{X}_2$

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2\ , \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

この式は一般に正しい

母分散既知 → Z 検定

母分散未知 → 等分散の検定

F 検定

等分散の場合 $\sigma_1^2 = \sigma_2^2$ → T 検定

等分散でない $\sigma_1^2 \neq \sigma_2^2$ → 高度な議論

Python を使ってみる

➢ 標本平均の差の分布

・2つの正規母集団 $N(m_1, s_1^2)$, $N(m_2, s_2^2)$ を扱う.

・大きさ n_1, n_2 の標本による標本平均の差 $D = \bar{X}_1 - \bar{X}_2$ の分布を調べる.

・仮説 $m_1 = m_2$ の下で同様の差 DH の分布を調べる.

・第 2 種誤り確率を確認する.

[TwoPopulations - Jupyter Notebook.pdf]

Lecture 10

母集団の比較

おわり