

統計学 01

早稲田大学政治経済学部
第16回
西郷 浩

1

本日の目標

- 母比率の検定 (p. 250)
- 母平均の差の検定 (pp. 242-244)
- 母比率の差の検定 (教科書に記載なし)
- 小標本における母平均の差の検定 (2006年6月30日追加)

2

母比率の検定 (1)

■ 右側検定

□ 仮説:

$$\begin{cases} H_0: p = p_0 \\ H_1: p = p_1 (> p_0) \end{cases} \quad \begin{cases} H_0: p = p_0 \\ H_1: p > p_0 \end{cases} \quad \begin{cases} H_0: p \leq p_0 \\ H_1: p > p_0 \end{cases}$$

□ 棄却域 D (有意水準 α):

$$D = \{ \hat{p} \mid \hat{p} > p_0 + z_\alpha \sqrt{p_0(1-p_0)/n} \}$$

$z_\alpha = (\text{標準正規分布の上側 } 100\alpha\% \text{点})$

3

母比率の検定 (2)

■ 左側検定

□ 仮説:

$$\begin{cases} H_0: p = p_0 \\ H_1: p = p_1 (< p_0) \end{cases} \quad \begin{cases} H_0: p = p_0 \\ H_1: p < p_0 \end{cases} \quad \begin{cases} H_0: p \geq p_0 \\ H_1: p < p_0 \end{cases}$$

□ 棄却域 D (有意水準 α):

$$D = \{ \hat{p} \mid \hat{p} < p_0 - z_\alpha \sqrt{p_0(1-p_0)/n} \}$$

$z_\alpha = (\text{標準正規分布の上側 } 100\alpha\% \text{点})$

4

母比率の検定(3)

■ 両側検定

□ 仮説: $H_0: p = p_0, H_1: p \neq p_0$

□ 棄却域 D (有意水準 α):

$$D = \{\hat{p} \mid \hat{p} < l, \text{ or } \hat{p} > u\}$$

$$l = p_0 - z_{\alpha/2} \sqrt{p_0(1-p_0)/n}$$

$$u = p_0 + z_{\alpha/2} \sqrt{p_0(1-p_0)/n}$$

$z_{\alpha/2}$ = (標準正規分布の上側 $100(\alpha/2)\%$ 点)

5

母比率の検定(4)

■ 例1(ホーエル p. 170)

□ メンデルの法則:

■ ある種のエンドウを交配させると $Y:G = 3:1$

□ 実験の結果(観察値):

■ $Y:G = 176:48 = 3.67:1$

■ 試行回数 $n = 176 + 48 = 224$

□ メンデルの法則は成立しないといえるか?

6

母比率の検定(5)

□ 解答:

■ 仮説の設定:

ただし、 $p = (Y \text{ が } \text{出る確率})$

■ 棄却域の決定(有意水準0.05):

$$D = \{\hat{p} \mid \hat{p} < l, \text{ or } \hat{p} > u\}$$

$$l = 3/4 - 1.96 \sqrt{(3/4)(1-3/4)/224} = 0.693$$

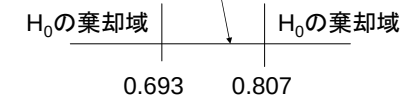
$$u = 3/4 + 1.96 \sqrt{(3/4)(1-3/4)/224} = 0.807$$

7

母比率の検定(6)

■ 検定統計量の観察:

$$\hat{p}_{obs} = 176/224 = 0.79$$



■ 検定の結果: H_0 は棄却されない。
(メンデルの法則は現実と矛盾しない。)

8

母比率の検定(7)

■ 例3(ホーエル p. 172):

□ 管理限界:

$$p \pm 3\sigma_{\hat{p}} \quad (\sigma_{\hat{p}}^2 = E[(\hat{p} - p)^2])$$

□ 信頼できる(n の大きい)標本から、管理限界を推定する。

□ 製造工程において、 $\hat{p}_{obs}$ が管理限界を超えるか否かをチェックする。

9

母比率の検定(8)

□ ミスタイプの管理:

■ タイプを練習している数名の学生を調べたところ、20,000字のうち800字がミスタイプだった。

■ 1時間に約600字を打つ学生がいたとして、1時間以内にタイプミスする字の割合の管理図を作る。

□ 解答:

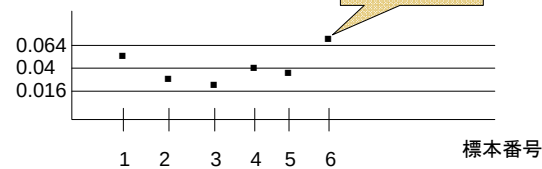
■ 真のミスタイプの割合 $p \doteq 800/20000 = 0.04$

■ 管理限界: $0.04 \pm 3(0.04 \times 0.96/600)^{1/2} = 0.016, 0.064$

10

母比率の検定(9)

■ 管理図



■ 管理図は、標本平均についても作ることができる(例4 ホーエル pp. 168-169)

11

母平均の差の検定 (1)

■ 例(ホーエル p. 172)

□ ふたつの銘柄の電球

■ 電球A: 平均寿命 μ_1 ; 分散 σ_1^2 (いずれの母数も未知)

■ 電球B: 平均寿命 μ_2 ; 分散 σ_2^2 (いずれの母数も未知)

□ μ_1 と μ_2 とに差があるが否かを調べたい。

12

母平均の差の検定(2)

□ 検定の方法

- m 個の電球Aを無作為抽出する。

標本平均 $\bar{X}$; 標本分散 s_1^2

- n 個の電球Bを無作為抽出する。

標本平均 $\bar{Y}$; 標本分散 s_2^2

(ふたつの標本平均が独立になることに注意)

- このとき:

$$\bar{X} - \bar{Y} \xrightarrow{D} N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}\right)$$

13

母平均の差の検定(3)

- このことから、たとえば「平均値に差があるかどうか」を検定するには:

- 仮説の設定:
$$\begin{cases} H_0: \mu_1 - \mu_2 = 0 \Leftrightarrow \mu_1 = \mu_2 \\ H_1: \mu_1 - \mu_2 \neq 0 \Leftrightarrow \mu_1 \neq \mu_2 \end{cases}$$

- 棄却域: 有意水準 α

$$D = \{\bar{x} - \bar{y} \mid \bar{x} - \bar{y} < l, \text{ or } \bar{x} - \bar{y} > u\}$$

$$l = 0 - z_{\alpha/2} \sqrt{\sigma_1^2/m + \sigma_2^2/n}$$

$$u = 0 + z_{\alpha/2} \sqrt{\sigma_1^2/m + \sigma_2^2/n}$$

$$z_{\alpha/2} = (N(0,1) \text{ の上側 } 100(\alpha/2)\% \text{ 点})$$

14

母平均の差の検定(4)

- 未知の母分散: σ_1^2, σ_2^2

- m と n がともに (25より) 大きいときには、標本分散が母分散に近いと考えて、それらを代入する。

$$\sigma_1^2 \doteq s_1^2, \sigma_2^2 \doteq s_2^2$$

- 検定統計量を観察して、帰無仮説 H_0 の採否を判定する。

15

母平均の差の検定(4)

□ 設問

- 電球Aの無作為標本 (大きさ $m=100$)

- 標本平均 1160、標本分散 90^2

- 電球Bの無作為標本 (大きさ $n=100$)

- 標本平均 1140、標本分散 80^2

□ 解答

- 仮説の設定:

- $H_0: \mu_1 = \mu_2$

- $H_1: \mu_1 \neq \mu_2$

16

母平均の差の検定(5)

- 棄却域の決定: 有意水準 0.05

$$D = \{\bar{x}_1 - \bar{x}_2 \mid \bar{x}_1 - \bar{x}_2 < l, \text{ or } \bar{x}_1 - \bar{x}_2 > u\}$$

$$l = 0 - 1.96 \sqrt{\frac{90^2}{100} + \frac{80^2}{100}} = -24, u = 24$$

$$\sigma_1^2 \doteq s_1^2 = 90^2, \sigma_2^2 \doteq s_2^2 = 80^2$$

- 観察された検定統計量:

$$(\bar{x}_1 - \bar{x}_2)_{obs} = 1160 - 1140 = 20 \notin D$$

- H_0 は棄却されない(平均寿命に差はない)。

17

母平均の差の検定(6)

- 片側検定 ($H_0: \mu_1 \leq \mu_2$; or $H_0: \mu_1 \geq \mu_2$) も、同様の考え方で実行できる。

- 「有意」= significant

- 差が有意である(有意差がある):

- $H_0: \mu_1 = \mu_2$ が棄却される。

- 結果が有意である:

- 観察された検定統計量 $\in$ 棄却域 D

- 標本の大きさ n_1, n_2 が大きいとき。

- 後述(古典的な難問として有名)

18

母比率の差の検定 (1)

- 例1(ホーエル p.176)

- 2種の船酔い予防薬: A, B

- 船員400人を無作為に二分する。

- グループ1(200人): Aを服用

- グループ2(200人): Bを服用

- 大しけのあとで、船酔いにならなかった者:

- グループ1: 152人

- グループ2: 132人

19

母比率の差の検定(2)

- AとBとで予防効果に差があるといえるか?

- 母数の定義:

- p_1 = Aを服用したときに船酔いにならない確率
(or たくさんのA服用者のうち船酔いしない者の比率)

- p_2 = Bを服用したときに船酔いにならない確率
(or たくさんのB服用者のうち船酔いしない者の比率)

- 調べるべきは、 p_1 と p_2 とに差があるか否か。

20

母比率の差の検定(3)

□ 検定の方法

- 無作為抽出した n_1 人にAを服用させ、船酔いしなかった人の(標本における)割合: $\hat{p}_1$
- 無作為抽出した n_2 人にBを服用させ、船酔いしなかった人の(標本における)割合: $\hat{p}_2$
- このとき、 n_1 と n_2 がある程度おおきければ:

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

21

母比率の差の検定(4)

- 母集団における比率 p_1, p_2 を指定すれば、バラツキの方は自動的に定まる!
- このことから、「 p_1 と p_2 とに差がない」という仮説を検定するには:

□ 帰無仮説の設定:

$$\begin{cases} H_0: p_1 - p_2 = 0 \Leftrightarrow p_1 = p_2 \\ H_1: p_1 - p_2 \neq 0 \Leftrightarrow p_1 \neq p_2 \end{cases}$$

□ 棄却域の設定: 有意水準 α

$$D = \{\hat{p}_1 - \hat{p}_2 \mid \hat{p}_1 - \hat{p}_2 < l, \text{ or } \hat{p}_1 - \hat{p}_2 > u\}$$

22

母比率の差の検定(5)

$$l = 0 - z_{\alpha/2} \sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}}$$

$\swarrow \quad \searrow$
 $n_1 \quad n_2$

$$= -z_{\alpha/2} \sqrt{p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

$p_1=p_2$ なので、
 $p_1=p_2=p$ とする。

$$u = z_{\alpha/2} \sqrt{p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

$z_{\alpha/2}$ = (標準正規分布の上側100($\alpha/2$)%点)

23

母比率の差の検定(6)

- $H_0: p_1 = p_2 (=p)$ のもとでの p の推定:

$\hat{p}$ = (グループ全体1, 2で船酔いしなかった者の比率)

理由:

H_0 のもとでは、AとBの効果に差がないから、グループ1とグループ2とは、予防薬の効果に関して同質的になる。

- $\hat{p}$ を p に代入して、棄却域を決定する。

□ 検定統計量の観察

□ H_0 の採否の判定

24

母比率の差の検定(7)

□ 例1(ホーエル p. 176)の解答:

■ 仮説の設定:

□ $H_0: p_1 = p_2$

□ $H_1: p_1 \neq p_2$

■ 棄却域の決定: 有意水準 0.05

$$\hat{p} = \frac{152 + 132}{200 + 200} = 0.71$$

$$D = \{\hat{p}_1 - \hat{p}_2 \mid \hat{p}_1 - \hat{p}_2 < l, \text{ or } \hat{p}_1 - \hat{p}_2 > u\}$$

$$l = -1.96\sqrt{0.71(1-0.71)(1/200+1/200)} = -0.09$$

$$u = 0.09$$

25

母比率の差の検定(8)

■ 検定統計量の観察:

$$(\hat{p}_1 - \hat{p}_2)_{obs} = \frac{152}{200} - \frac{132}{200} = 0.10$$

■ 判定:

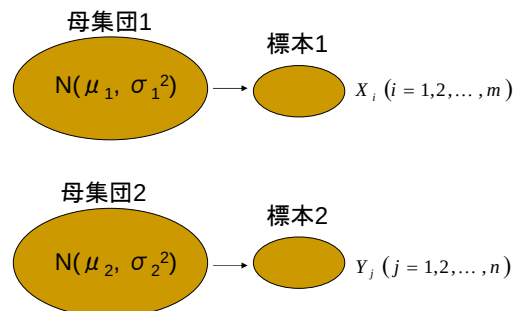
$$(\hat{p}_1 - \hat{p}_2)_{obs} \in D$$

したがって、 H_0 は棄却される(AとBとには効果に差がある)。

26

小標本における母平均の差の検定(1)

■ 二標本問題



27

小標本における母平均の差の検定(2)

このとき、

$$\bar{X} = \frac{1}{m} \sum_{i=1}^m X_i, \quad \bar{Y} = \frac{1}{n} \sum_{j=1}^n Y_j$$

とすると、 $\bar{X} \sim N(\mu_1, \sigma_1^2/m)$, $\bar{Y} \sim N(\mu_2, \sigma_2^2/n)$ かつ両者が独立になるので、

$$\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \sigma_1^2/m + \sigma_2^2/n)$$

または

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} \sim N(0, 1)$$

28

小標本における母平均の差の検定(3)

σ_1^2 と σ_2^2 とをそれぞれ

$$s_1^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2, \quad s_2^2 = \frac{1}{n-1} \sum_{j=1}^n (Y_j - \bar{Y})^2$$

でおきかえたとする、

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{s_1^2/m + s_2^2/n}} \sim ? \text{ (とても複雑な分布)}$$

Behrens–Fisherの問題として有名。

29

小標本における母平均の差の検定(4)

ただし、 $\sigma_1^2 = \sigma_2^2$ のときには解決されている。

$\sigma_1^2 = \sigma_2^2 = \sigma^2$ とおけば、

$$\sigma_1^2/m + \sigma_2^2/n = \sigma^2(1/m + 1/n)$$

$\sigma_1^2 = \sigma_2^2$ がたまたまのときの σ^2 の推定量として

$$\begin{aligned} s^2 &= \frac{\sum_{i=1}^m (X_i - \bar{X})^2 + \sum_{j=1}^n (Y_j - \bar{Y})^2}{m + n - 2} \\ &= \frac{(m-1)s_1^2 + (n-1)s_2^2}{(m-1) + (n-1)} \end{aligned}$$

を使う。

30

小標本における母平均の差の検定(5)

このとき、

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{s^2(1/m + 1/n)}} \sim t(m+n-2)$$

が示せる。とくに、 $H_0: \mu_1 - \mu_2 = 0$ が正しいとき

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

これを利用して、母数に依存しない棄却域を設定できる。

31

小標本における母平均の差の検定(6)

つまり、

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

とおくとき、

$$W = \{t : t < -t_{\alpha/2}(m+n-2), \text{ or } t > t_{\alpha/2}(m+n-2)\}$$

ただし、 $t_{\alpha/2}(m+n-2)$ は自由度 $m+n-2$ の t 分布の

上側 $100(\alpha/2)\%$ 点

とすれば、 W が両側検定の棄却域となる。

32

小標本における母平均の差の検定(7)

まとめと：

$X_i (i = 1, 2, \dots, m)$ が $N(\mu_1, \sigma_1^2)$ からの無作為標本

$Y_j (j = 1, 2, \dots, n)$ が $N(\mu_2, \sigma_2^2)$ からの無作為標本

のとき、

$$\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \neq 0 \end{cases}$$

を有意水準 0.05 で検定するには、

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

として (つぎに続く)

33

小標本における母平均の差の検定(8)

H_0 の棄却域を

$$W = \{t : t < -t_{0.025}(m+n-2), \text{ or } t > t_{0.025}(m+n-2)\}$$

とする。したがって、

$$t_{obs} \in W \Rightarrow H_0 \text{を棄却する。}$$

$$t_{obs} \notin W \Rightarrow H_0 \text{を棄却しない。}$$

W の代わりに以下の D を H_0 の棄却域としてもよい。

$$D = \{\bar{x} - \bar{y} : \bar{x} - \bar{y} < l, \text{ or } \bar{x} - \bar{y} > u\}$$

$$\text{ただし、} l = -t_{0.025}(m+n-2)\sqrt{s^2(1/m+1/n)}$$

$$u = t_{0.025}(m+n-2)\sqrt{s^2(1/m+1/n)}$$

34

小標本における母平均の差の検定(9)

■ 問題12.2 (p. 252) i)

- 男女の賃金格差の有無を調べるために、同年齢・同学歴・同職種の男女それぞれ10人ずつを無作為抽出した。
- 賃金を調べたところ、以下の表のようになった。
- 「男女間で賃金に格差がある」といえるか。

表1: 無作為抽出された男女の賃金(万円)[架空例]

男	15.4	18.3	16.5	17.4	18.9	17.2	15.0	15.7	17.9	16.5
女	14.2	15.9	16.0	14.0	17.0	13.8	15.2	14.5	15.0	14.4

35

小標本における母平均の差の検定(10)

まず、

男子の賃金 $X_i (i = 1, 2, \dots, m)$ が $N(\mu_1, \sigma_1^2)$ からの無作為標本

女子の賃金 $Y_j (j = 1, 2, \dots, n)$ が $N(\mu_2, \sigma_2^2)$ からの無作為標本

とみなす。

$$\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \neq 0 \end{cases}$$

を有意水準0.05で検定する。 H_0 の棄却域は、

$$t = (\bar{x} - \bar{y}) / \sqrt{s^2(1/m + 1/n)}$$

として、

$$W = \{t : t < -2.101, \text{ or } t > 2.101\}$$

36

小標本における母平均の差の検定(11)

観察値にもとづいて統計量を計算すると、

$$\bar{x} = 16.88, s_1^2 = 1.66, \bar{y} = 15.00, s_2^2 = 1.06$$

$$s^2 = 1.36$$

これらから、

$$t_{obs} = \frac{16.88 - 15.00}{\sqrt{1.36(1/10 + 1/10)}} = 3.61$$

したがって、

$$t_{obs} \in W \Rightarrow H_0 \text{を棄却する。}$$