

平成22年度 統計工学

平成22年度
統計工学 講義資料(9)

～統計的推論:仮説検定III～

(C) Sadami SUZUKI

平成22年度 統計工学

母集団が正規分布に従うとの仮定の下に…

・平均値の検定

- 母分散が既知の場合…正規分布に基づく検定(基準化)
- 母分散が未知の場合…t分布に基づく検定(スチューデント化)
- 2つの母集団平均に関する比較(対応あり)

(C) Sadami SUZUKI

平成22年度 統計工学

2つの母集団比較:t検定

店舗Xと店舗Yにおける顧客満足度の比較

店舗XとYとでは どちらがよい顧客からの 支持を得ているだろうか?

(C) Sadami SUZUKI

平成22年度 統計工学

2つの母集団比較:t検定(対応なし)

前回の復習…

- 店舗Xおよび店舗Yの主婦層を対象に行った顧客満足度調査により表1に示すデータが得られたとする
- これら2店舗の平均満足度には差があると言えるか

| 番号 | 店舗X | 店舗Y |
|----|-----|-----|
| 1 | 93 | 84 |
| 2 | 75 | 97 |
| 3 | 72 | 89 |
| 4 | 79 | 95 |
| 5 | 80 | 85 |
| 6 | 82 | 81 |
| 7 | 72 | 77 |
| 8 | 78 | 82 |
| 9 | 89 | 75 |

【等分散性の確認】

$$\frac{\hat{\sigma}_X^2}{\hat{\sigma}_Y^2} = \frac{56.25}{51.50} = 1.09 < F_{\alpha}^*(0.05) = 3.44$$

仮説は有意水準5%で棄却できない

【等分散性を仮定した t検定】

$$t_0 = \frac{\bar{X} - \bar{Y}}{\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right) \hat{\sigma}^2}} = \frac{412 - 450}{\sqrt{\frac{1}{9} + \frac{1}{9}} \times 53.875} = 1.445 < t_{\alpha}(0.05) = 2.120$$

仮説は有意水準5%で棄却できない

両店舗間の満足度に統計的有意差があるとは言えない

(C) Sadami SUZUKI

平成22年度 統計工学

2つの母集団比較:t検定(対応なし)

前回扱った対応のないt検定では…

店舗Xの顧客と店舗Yの顧客における

・満足度の平均値の差

に着目した検定を行った

(C) Sadami SUZUKI

平成22年度 統計工学

2つの母集団比較:t検定(対応なし)

検定仮説

$$H_0: \mu_X = \mu_Y$$

Step2:データ収集および基本統計量の計算

- 仮説に応じたデータの収集(母集団からの標本抽出)
- 基本統計量の計算

店舗X, 店舗Yそれぞれの顧客(主婦層)からの無作為標本抽出が条件

店舗Xの評価者と店舗Yの評価者が異なる(2つの母集団が独立)

(C) Sadami SUZUKI

平成22年度 統計工学

平均値の比較:t検定(対応あり)

今回扱う**対応のある**t検定は…

店舗Xと店舗Yにおける

満足度の差の平均値

に着目した検定手法です

(C) Sudami SUZUKI

平成22年度 統計工学

「対応」の意味

前回の例において、店舗Xの評価を行った顧客Aさんと店舗Yの評価を行った顧客Bさんの点数の差は意味をなさない

同じ顧客Aさんに店舗Xと店舗Yの評価をしてもらう場合は？

| 番号 | x_i | y_i | $x_i - y_i$ |
|----|-------|-------|-------------|
| 1 | 93 | 84 | 9 |
| 2 | 75 | 97 | -22 |
| 3 | 72 | 89 | -17 |
| 4 | 79 | 95 | -16 |
| 5 | 80 | 85 | -5 |
| 6 | 82 | 81 | 1 |
| 7 | 72 | 77 | -5 |
| 8 | 78 | 82 | -4 |
| 9 | 89 | 75 | 14 |
| 合計 | 720 | 765 | -45 |

直接比較することは出来ない

| 番号 | x_i | y_i | $x_i - y_i$ |
|----|-------|-------|-------------|
| 1 | 72 | 75 | -3 |
| 2 | 80 | 84 | -4 |
| 3 | 82 | 87 | -5 |
| 4 | 78 | 83 | -5 |
| 5 | 72 | 82 | -10 |
| 6 | 89 | 94 | -5 |
| 7 | 79 | 83 | -4 |
| 8 | 75 | 80 | -5 |
| 9 | 93 | 88 | 5 |
| 合計 | 720 | 756 | -36 |

直接比較可能

(C) Sudami SUZUKI

平成22年度 統計工学

2つの母集団比較:t検定(対応あり)

Step2:データ収集および基本統計量の計算

- 仮説に応じたデータの収集(母集団からの標本抽出)
- 基本統計量の計算

「1つの母集団から標本の無作為抽出」に基づいて比較を行う

顧客(主婦層)に店舗Xと店舗Yの両方を評価をもらい、直接比較可能な満足度の差について考える

(C) Sudami SUZUKI

平成22年度 統計工学

2つの母集団比較:t検定(対応あり)

Step1:仮説の設定

- 問題意識の明確化
- 考察の対象となる母集団特性値に関する仮説の設定

店舗Xの店長は最近出店してきた競合店Yと比較して主婦層のウケが悪いように感じている…

「うちの顧客満足度は店舗Yに比べてどうだろうか？」

帰無仮説 $H_0: \mu = 0$

対立仮説 $H_1: \mu < 0$ …片側検定
 $H_1: \mu > 0$ …片側検定
 $H_1: \mu \neq 0$ …両側検定

「顧客(主婦層)に店舗Xと店舗Yの両方を評価してもらって、満足度に差があるか？」

(C) Sudami SUZUKI

平成22年度 統計工学

2つの母集団比較:t検定(対応あり)

Step2:データ収集および基本統計量の計算

- 仮説に応じたデータの収集(母集団からの標本抽出)
- 基本統計量の計算

検定仮説 $H_0: \mu = 0$

【店舗Xに対する評価】観測値: x_1, x_2, \dots, x_n
 確率変数: X_1, X_2, \dots, X_n

【店舗Yに対する評価】観測値: y_1, y_2, \dots, y_n
 確率変数: Y_1, Y_2, \dots, Y_n

満足度の差に関する観測値: $d_i = x_i - y_i$
 確率変数: $D_i = X_i - Y_i$

(C) Sudami SUZUKI

平成22年度 統計工学

2つの母集団比較:t検定(対応あり)

Step3:検定に用いる統計量の選択とその分布に関する検討

- 考察の対象: **2つの店舗に関する満足度の差**
- 検定量: 2つの店舗に関する満足度の差に関する最良不偏推定量としての標本平均をベースに考える

検定仮説 $H_0: \mu = 0$

$$T = \frac{\bar{D} - \mu}{\hat{\sigma} / \sqrt{n}} = \frac{\bar{D} - \mu}{\sqrt{\frac{1}{n} \times \frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2}}$$

自由度 $n-1$ の t 分布に従う

(C) Sudami SUZUKI

平成22年度 統計工学

2つの母集団比較:t検定(対応あり)

検定仮説

Step4:検定量を用いて仮説が正しいか否かを判定する

帰無仮説が成り立つとき

$$T_0 = \frac{\bar{D}}{\hat{\sigma} / \sqrt{n}}$$

自由度 $n-1$ の t 分布に従う

得られた標本を基に検定量の計算 $t_0 = \frac{\bar{d}}{\hat{\sigma} / \sqrt{n}}$

$|t_0| > t_{n-1}(\alpha)$ となる確率が α

有意水準 α で仮説は棄却される

$H_0: \mu = 0$

(C) Sadami SUZUKI

平成22年度 統計工学

2つの母集団比較:t検定(対応あり)

【確認テスト1】

主婦層を対象に店舗X、店舗Yの両店舗を評価してもらった顧客満足度調査により表1に示すデータが得られたとする

これら2店舗の平均満足度には差があると言えるか

表1 両店舗の満足度

| 番号 | 店舗X | 店舗Y |
|----|-----|-----|
| 1 | 72 | 75 |
| 2 | 80 | 84 |
| 3 | 82 | 87 |
| 4 | 78 | 83 |
| 5 | 72 | 82 |
| 6 | 89 | 94 |
| 7 | 79 | 83 |
| 8 | 75 | 80 |
| 9 | 93 | 88 |

(C) Sadami SUZUKI

平成22年度 統計工学

2つの母集団比較:t検定(対応あり)

【確認テスト1】

主婦層を対象に店舗X、店舗Yの両店舗を評価してもらった顧客満足度調査により表1に示すデータが得られたとする

これら2店舗の平均満足度には差があると言えるか

表1 両店舗の満足度

| 番号 | 店舗X | 店舗Y |
|----|-----|-----|
| 1 | 72 | 75 |
| 2 | 80 | 84 |
| 3 | 82 | 87 |
| 4 | 78 | 83 |
| 5 | 72 | 82 |
| 6 | 89 | 94 |
| 7 | 79 | 83 |
| 8 | 75 | 80 |
| 9 | 93 | 88 |

<解答例>

| 番号 | x | y | (x-y) | (x-y) ² | (x-y) ³ |
|----|-----|-----|-------|--------------------|--------------------|
| 1 | 72 | 75 | -3 | 9 | -27 |
| 2 | 80 | 84 | -4 | 16 | -64 |
| 3 | 82 | 87 | -5 | 25 | -125 |
| 4 | 78 | 83 | -5 | 25 | -125 |
| 5 | 72 | 82 | -10 | 100 | -1000 |
| 6 | 89 | 94 | -5 | 25 | -125 |
| 7 | 79 | 83 | -4 | 16 | -64 |
| 8 | 75 | 80 | -5 | 25 | -125 |
| 9 | 93 | 88 | 5 | 25 | 125 |
| 合計 | 720 | 756 | -36 | 360 | -3240 |

$\bar{d} = \frac{-36}{9} = -4.0$

$s^2 = \frac{360}{9-1} = 45.0$

$s = 3.905$

$|t_0| = \frac{-4.0}{3.905/\sqrt{9}} = 3.073 > t_4(0.05) = 2.306$

仮説は有意水準5%で棄却される

両店舗間の満足度は5%有意で統計的な差がある

(C) Sadami SUZUKI

平成22年度 統計工学

平均値の比較

<例>

- 異なる工作機械によって製造された部品の品質比較
- 各種文化指標の国際比較

対応のないt検定

- 既存手法と改善案との優劣比較
- 投薬治療の効果測定

対応のあるt検定

(C) Sadami SUZUKI

平成22年度 統計工学

バラツキに関する検定

・ χ^2 適合度検定

- 観測度数(出現率), 分割表に基づく検定手法

(C) Sadami SUZUKI

平成22年度 統計工学

観測度数に基づく比較: χ^2 適合度検定

以下は、無作為抽出された100人を対象に行ったインフルエンザの感染と予防接種の有無についての調査結果であるとする。

| | インフルエンザ | | 計 |
|-------|---------|---------|-----|
| | 感染した | 感染しなかった | |
| 予防接種 | 5 | 35 | 40 |
| しなかった | 15 | 45 | 60 |
| 計 | 20 | 80 | 100 |

感染者の割合:

【予防接種をした場合】

$$\frac{5}{40} = 12.5(\%)$$

【予防接種をしなかった場合】

$$\frac{15}{60} = 25(\%)$$

予防接種は有効であると言えるだろうか?

予防接種の有無によってインフルエンザに感染する割合は変わると言えるか?

(C) Sadami SUZUKI

