講義にて配布した資料の抜粋・縮小版です

平成22年度
統計工学　講義資料(7)

～統計的推論：仮説検定～

(C) Sadami SUZUKI

統計的仮説検定の枠組み

統計的仮説検定

母集団分布とある基準分布との違い，あるいは2つ以上の母集団分布の違いを，それぞれの母集団からの無作為抽出された標本に基づき客観的に認識するための方法論

「違い」があることを積極的に主張するための手段

(C) Sadami SUZUKI

仮説検定の判断枠組と背理法

背理法

- 論証したい事象を否定する仮定を立てる
- 仮定から導かれる事象に矛盾が生じることを示す
- 仮定が誤りであることを示す

検定の論理

・「確率的」な背理法と位置づけられる

「仮に違いがまったくないとしたときに，標本が観測された値をとることは非常に稀である」　…という判断枠組みに基づく

(C) Sadami SUZUKI

帰無仮説と対立仮説

平均値の検定・・・ $\mu < \mu_0,\quad \mu = \mu_0,\quad \mu > \mu_0$

帰無仮説

- 否定するために設定する命題　$H_0 : \mu = \mu_0$

対立仮説

- 帰無仮説を否定することで対立的に成り立つと考える命題
  - 片側対立仮説
    $H_1 : \mu < \mu_0$
    $H_1 : \mu > \mu_0$
  - 両側対立仮説
    $H_1 : \mu \neq \mu_0$

(C) Sadami SUZUKI

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(対応なし，対応あり)

(C) Sadami SUZUKI

平均値の検定

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の店長は系列全店の顧客満足度調査の結果(平均72点)に比べて，自身の店舗の満足度が気にかかっている・・・

「ウチの顧客満足度は全国平均に比べてどうだろうか？」

検定仮説

(C) Sadami SUZUKI

講義にて配布した資料の抜粋・縮小版です

平均値の検定

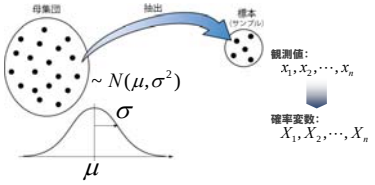
Step2:データ収集および基本統計量の計算

- 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

$$\mu = \mu_0 = 72$$

番号	店舗X
1	93
2	75
3	72
4	79
5	80
6	82
7	72
8	78
9	89



(C) Sadami SUZUKI

平均値の検定

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:**母集団平均値**
- 検定量:母平均の**最良不偏推定量**としての**標本平均**をベースに考える

$$\hat{\mu} = \bar{X} = \frac{\sum X_i}{n}$$

- 分布:基準化に基づく**正規分布**, スチューデント化に基づく ***t* 分布**

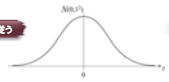
検定仮説

$$\mu = \mu_0 = 72$$

標準化

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

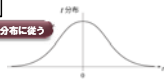
標準正規分布に従う



スチューデント化

$$T = \frac{\bar{X} - \mu}{\hat{\sigma} / \sqrt{n}}$$

自由度 *n*-1 の *t* 分布に従う



(C) Sadami SUZUKI

平均値の検定

Step4:検定量を用いて仮説が正しいか否かを判定する

- 検定量の値について境界値を設定し, 仮説の成立について判定する
- **仮説の棄却**(否定)
  - 仮説が正しいとすると, その検定量の値が得られる確率が非常に小さいと考えられる領域
  (棄却域)では, 仮説を棄却する
- **仮説の支持**(採択)
  - 仮説が正しいとして, その検定量の値が得られる確率が小さくはない領域**(採択域)**
  では, 仮説を支持する

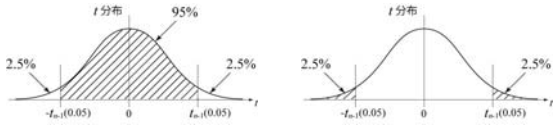
仮説

$$\mu = \mu_0 = 72$$

(C) Sadami SUZUKI

推定と検定は表裏一体

- 仮説の成立を判定するための境界値
- **有意水準 α** …5%, 1%が用いられることが多い
- 区間推定における**95%信頼区間から外れる確率**が**有意水準5%**
(両側検定)



(C) Sadami SUZUKI

平均値の検定

Step4:検定量を用いて仮説が正しいか否かを判定する

- 95%信頼区間は有意水準5%の採択域

$$\Pr\left[\left|\frac{\bar{x} - \mu}{\hat{\sigma} / \sqrt{n}}\right| \le t_{n-1}(0.05)\right] = 0.95 \quad \Rightarrow \quad \Pr\left[\left|\frac{\bar{x} - \mu}{\hat{\sigma} / \sqrt{n}}\right| > t_{n-1}(0.05)\right] = 0.05$$

- 仮説として設定した値を基準に考えて…

$$\left|\frac{\bar{x} - \mu_0}{\hat{\sigma} / \sqrt{n}}\right| > t_{n-1}(\alpha)$$ となる確率がα

有意水準αで仮説は棄却される

100(1−α)%信頼区間を活用

仮説

$$\mu = \mu_0 = 72$$

(C) Sadami SUZUKI

検定の流れと2種類の誤り

		仮説が正しいとき	仮説が誤りの時
仮説検定結果	仮説を採択(支持)	正	第2種の誤り β
	仮説を否定(棄却)	第1種の誤り α	正

消費者危険
(Consumer's risk)

生産者危険
(Producer's risk)

検出力

$1-\beta$ …誤った仮説を採択しない(否定できる)確率

(C) Sadami SUZUKI