第6回　5月7日の授業内容

- §2.4　分布の量的記述
  - §2.4.1　分布の中心を示す代表値
    - モード
    - メディアン
    - 算術平均
    - その他
  - §2.4.2　分布の広がりを示す代表値
    - レインジ
    - 平均偏差
    - 分散と標準偏差

5/7/03 1

§2.4.1 分布の中心を示す代表値(1)

◈モード（最頻値） mode
- 標本データの「最頻値」を示す代表値
- 連続データの場合、モードは最も頻度が高い区間（階級）“most popular class”の区間代表値（階級値）とする。

5/7/03 2

§2.4.1 分布の中心を示す代表値(2)

◈メディアン（中央値） median
- 累積相対頻度が50%に達したときの値。
- 標本データを昇順に並べたときの真中の個体値。
- 標本数が奇数のとき　(n+1)/2 番目の個体
  標本数が偶数のとき　n/2番目とn/2+1番目
  の個体の算術平均

5/7/03 3

§2.4.1 分布の中心を示す代表値(3)

◈Note: **分位点** percentile

- 標本データを昇順に並べて、ある観測値を含めて、それよりも小さい標本の数が全体のα%（0≦α≦1）であるとき、その観測値をデータ分布のα%分位点という。
- αとして、25、50、75が用いられ、それぞれ「第1四分位点」、「メディアン」、「第3四分位点」と呼ばれる。

5/7/03 4

§2.4.1 分布の中心を示す代表値(4)

◈**算術平均**（または単に、平均）
(arithmetic) mean

- データの分布の重心を示す統計量

$$平均 = \frac{\sum_{i=1}^{n} x_i}{n}$$

x_i は標本データ、nは標本数を示す。

5/7/03 5

§2.4.1 分布の中心を示す代表値(5)

◈**その他**(1)

- 切り落とし平均(trimmed mean)
  異常値、外れ値の影響を除去するために、最大値、最小値とその周囲の観測値を除いて算術平均をとったもの
- 幾何平均(geometric mean)
  平均成長率（増加率）

5/7/03 6

§2.4.1 分布の中心を示す代表値(6)

◈**その他**(2)

- **加重平均**(weighted mean)
  観測値にウエイトをかけて平均をとる。
  但し、ウエイトの和は1。算術平均はスペシャルケース。

$$加重平均 = \sum_{i=1}^{n} c_i x_i, \quad 但し \sum_{i=1}^{n} c_i = 1$$

5/7/03 7

§2.4.2 分布の広がりを示す代表値(1)

◈**レインジ(範囲)** range

- **レインジ≡最大値－最小値**
  異常値に敏感なのが欠点

◈**四分位レインジ** interquartile range

- **四分位レインジ≡第3四分位点－第1四分位点**
  レインジや標準偏差よりは頑強

5/7/03 8

§2.4.2 分布の広がりを示す代表値(2)

◈**平均偏差**

- **平均的に(算術)平均から乖離している距離**

$$平均偏差 \equiv \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

5/7/03 9

§2.4.2 分布の広がりを示す代表値(3)

◈分散と標準偏差
- 分散 variance
  - 偏差2乗和の平均で分布の広がりを表したもの。

$$\text{分散}\ \sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

  - 分散の利点
    - 絶対偏差より数学的な扱いが楽
    - 平均からの距離を誇張する機能⇒異常値に弱い

5/7/03 10

§2.4.2 分布の広がりを示す代表値(4)

- 標準偏差 standard deviation
  - 分散だと（単位）^{2}になってしまうので、他の分布の広がりを示す代表値と単位を合わせるために（比較可能にするため）、分散の平方根をとる。⇒ 標準偏差

$$\text{標準偏差}\ \sigma = \sqrt{\sigma^2}$$

5/7/03 11

データ例（再録）

◈O大学下宿生の通学時間（分）
標本数１９

40, 15, 60, 15, 45, 10, 15, 15, 18, 20, 47, 10, 5, 15, 20, 60, 15, 25, 10

5/7/03 12