

統計学入門 第5回

早稲田大学政治経済学部
西郷 浩

1

本日の目標

- 度数分布表からの近似計算
 - ヒストグラムの作成
 - 算術平均、分散の計算
 - 中央値、四分位偏差の計算
 - ローレンツ曲線の描画、ジニ係数の計算
- PC実習

2

度数分布表による統計の提供(1)

- 公表されている統計(の多く)
 - 原データ x_i : 利用不可
 - 度数分布表: 利用可
 - 例: 総務省統計研修所(2011)
『第61回日本統計年鑑』
 - 表18-17 敷地面積別一戸建住宅数(借家、2008年)

3

度数分布表による統計の提供(2)

表1: 敷地面積別一戸建住宅数(借家、2008年)

階級	下限(m^2 以上)	上限(m^2 未満)	度数(1000戸)
1	-	50	214
2	50	75	427
3	75	100	351
4	100	150	383
5	150	200	245
6	200	300	177
7	300	500	89
8	500	-	35
		合計	1,921

資料: 総務省統計研修所編(2011)『第61回日本統計年鑑』表18-17

4

ヒストグラムの作成(1)

■ 度数分布表からのヒストグラムの作成

■ 注意点

- 異なる階級幅が混在している。
 - 原則: 柱の面積が度数に比例するように描く。
- ヒストグラムの縦軸を密度にすると安全である。
 - 密度 = 相対度数 ÷ 階級幅
- 開放間隔の階級が存在
 - 開放間隔: 下限または上限が明示されていないこと。
 - 自分で想定してヒストグラムを作成する。

5

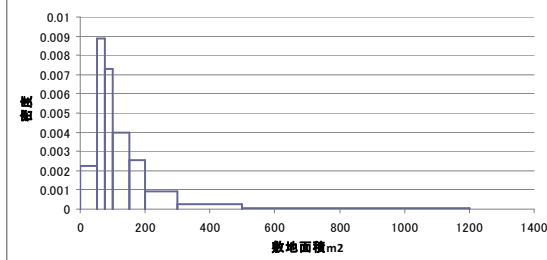
ヒストグラムの作成(2)

- 開放間隔の階級についての想定
 - 第1階級の下限: 0m^2
 - 第8階級の上限: 1200m^2
- 大胆な仮定を置くことになるので、明示する必要あり。
- 算術平均等の計算でも、第1階級の下限は 0m^2 、第8階級の上限: 1200m^2 とする。

6

ヒストグラムの作成(3)

図1: 一戸建敷地面積のヒストグラム



資料: 総務省統計研究所編(2011)『第61回日本統計年鑑』表18-17

7

算術平均の近似計算(1)

■ 階級内一定値の仮定

- ある階級に属する個体は、すべて階級値に等しい値を持つ。
 - 例: 第2階級について
 - 階級値: $(50+75)/2 = 62.5$
 - 度数: 427(千戸)
 - 第2階級に属する個体については、すべて $x_i = 62.5$ と考える。
 - その仮定のもとで:
 - 第2階級に属する個体の x_i の合計:
 $427(\text{千戸}) \times 62.5 = 26687.5 (\times 1000)$

8

算術平均近似計算(2)

仮定のもとで：

$$\begin{aligned}\bar{x} &= \frac{1}{1921 \times 1000} \{ 214 \times 1000 \times 25.0 + 427 \times 1000 \times 62.5 \\ &\quad + 351 \times 1000 \times 87.5 + 383 \times 1000 \times 125.0 \\ &\quad + 245 \times 1000 \times 175.0 + 177 \times 1000 \times 250.0 \\ &\quad + 89 \times 1000 \times 400 + 35 \times 1000 \times 850 \} \\ &\approx 137.0\end{aligned}$$

9

算術平均近似計算(3)

階級値の加重平均：

$$\begin{aligned}\bar{x} &= \frac{214}{1921} \times 25.0 + \frac{427}{1921} \times 62.5 + \frac{351}{1921} \times 87.5 \\ &\quad + \frac{383}{1921} \times 125.0 + \frac{245}{1921} \times 175.0 + \frac{177}{1921} \times 250.0 \\ &\quad + \frac{89}{1921} \times 400 + \frac{35}{1921} \times 850 \\ &\approx 137.0\end{aligned}$$

10

分散・標準偏差の近似計算(1)

- 階級内一定値の仮定
 - すべての個体の x の値が定まる。
 - 同じ仮定のもとで算術平均を近似する。
 - 計算済み。
 - 分散の公式どおりに計算して近似する。
 - 平均からの偏差の2乗の平均値。
 - 標準偏差＝分散の平方根

11

分散・標準偏差の近似計算(2)

$$\begin{aligned}S^2 &= \frac{1}{1921} \{ 214 \times (25.0 - 137.0)^2 + 427 \times (62.5 - 137.0)^2 \\ &\quad + 351 \times (87.5 - 137.0)^2 + 383 \times (125.0 - 137.0)^2 \\ &\quad + 245 \times (175.0 - 137.0)^2 + 177 \times (250.0 - 137.0)^2 \\ &\quad + 89 \times (400.0 - 137.0)^2 + 35 \times (850.0 - 137.0)^2 \} \\ &= 16935 \\ S &= 130\end{aligned}$$

12

中央値の近似計算(1)

■ 階級内一様分布の仮定

- ある階級の中に属する個体が、階級幅全体に一様に分布している。
 - たとえば、第2階級について：
 - 下限：50、上限：75、階級幅：25(=75-50)
 - 度数：427(千戸)
 - 下限(=50) から、階級幅の 1/5 まで(50+25/5 = 55)の幅に、度数の1/5(=427/5 千戸)が分布している。
 - 下限から階級幅の 2/5 まで(=60)の幅に、度数の 2/5(=427×2/5) が分布している、等々。

13

中央値の近似計算(2)

表2: 敷地面積別一戸建住宅数(借家、2008年、累積相対度数入り)

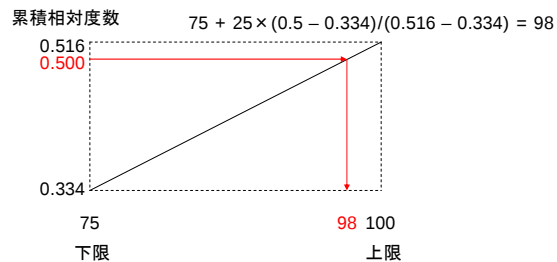
階級	下限(m ² 以上)	上限(m ² 未満)	度数(1000戸)	累積相対度数
1	-	50	214	0.111
2	50	75	427	0.334
3	75	100	351	0.516
4	100	150	383	0.716
5	150	200	245	0.843
6	200	300	177	0.935
7	300	500	89	0.982
8	500	-	35	1.000
合計			1,921	

資料：総務省統計研究所編(2011)『第61回日本統計年鑑』表18-17

14

中央値の近似計算(3)

図2: 第3階級における近似的な中央値の位置



15

四分位偏差の近似計算(1)

■ 階級内一様分布の仮定

- 四分位点も中央値と同じように計算する。

■ 第1四分位点 Q_1 :

- 表2から、第2階級にふくまれる。
 - $Q_1 = 50 + (75 - 50) \times (0.25 - 0.111) / (0.334 - 0.111) = 66.6$

■ 第3四分位点 Q_3 :

- 表2から、第5階級にふくまれる。
 - $Q_3 = 150 + (200 - 150) \times (0.75 - 0.716) / (0.843 - 0.716) = 163.4$

16

四分位偏差の近似計算(2)

- 四分位範囲
 - $Q_3 - Q_1 = 163.4 - 66.6 = 96.8$
- 四分位偏差
 - $d = (Q_3 - Q_1)/2 = 48.4$

17

ローレンツ曲線の作成 (1)

- 階級内一定値の仮定
 - ある階級に属する個体は、すべて階級値に等しい値を持つ。
 - その階級の中では、個体数の相対的な累積のスピードと、変数の相対的な累積量のスピードとが同じになる。
 - その階級の中では、ローレンツ曲線が直線になる。

18

ローレンツ曲線の作成 (2)

- ローレンツ曲線の作成手順
 - 横軸: 階級 j の累積相対度数
 - 縦軸: 階級 j の変数の相対的な累積量
 - 計算方法:
 - 度数 × 階級値で各階級の変数合計を近似する。
 - 算術平均の近似と同じ計算
 - 近似値を累積する。
 - 累積値を合計で除して相対的な累積量にする。

19

ローレンツ曲線の作成 (3)

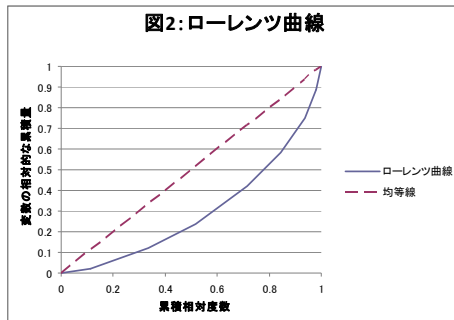
表3: 変数の相対的な累積量の計算

階級	下限 (m2以上)	上限 (m2未満)	度数 (1000戸)	累積相対 度数	階級値	変数合計	変数の相 対的な累 積量
1	0	50	214	0.111	25.0	5350.0	0.020
2	50	75	427	0.334	62.5	26687.5	0.122
3	75	100	351	0.516	87.5	30712.5	0.239
4	100	150	383	0.716	125.0	47875.0	0.420
5	150	200	245	0.843	175.0	42875.0	0.583
6	200	300	177	0.935	250.0	44250.0	0.752
7	300	500	89	0.982	400.0	35600.0	0.887
8	500	1200	35	1.000	850.0	29750.0	1.000
		合計	1921			263100.0	

資料: 総務省統計研究所編(2011)『第61回日本統計年鑑』表18-17

20

ローレンツ曲線の作成 (4)



資料: 総務省統計研究所編(2011)『第61回日本統計年鑑』表18-17

21

ジニ係数の計算 (1)

- ローレンツ曲線から、通常と同じように計算する。
 - 台形の面積の公式を利用して、ローレンツ曲線の下側の面積を計算する。
 - その2倍を1から減じる。
 - 借家の一戸建敷地面積については、
 - $GI = 0.41$

22

ジニ係数の計算 (2)

- 度数分布表からの近似計算の方が、もとのデータから計算したものよりも、不均等度が低くなりやすい。
 - ローレンツ曲線は均等線に近くなる。
 - ジニ係数は0に近くなる。
 - 階級内一定値の仮定(つまり、階級内の不均等が無視される)ため。

23

PC実習

- ヒストグラムの作成
- 算術平均・分散の近似計算
- 中央値・四分位偏差の近似計算
- ローレンツ曲線の近似、ジニ係数の近似計算

24