

第10回 6月27日の講義内容

- § 4. 予備的分析
 - 作表
 - 独立性検定、相関
- § 5. 調査例: 政府の統計調査

6/27/07

1

§ 4. 予備的分析

- 作表
 - 単純集計
 - クロス集計: ex.) 分割表
平均や標準偏差など統計量を見るだけでなく、分布の形状にも注意
 - Bi-modalのケース
 - 歪んだ分布のケース

6/27/07

2

§ 4. 予備的分析(2)

- 独立性の検定: カイ2乗検定
 - 関係性の有無をチェック
統計手法のみからは因果関係はわからない
- 相関係数
 - 相関の有無、程度をみる
検定を行う場合には、帰無仮説に注意!! $H_0: r=0$
 - 相関 = 二変数の線形関係の程度を表す尺度
相関係数からは非線形関係の存在は分らない

6/27/07

3

クロス集計の例

(上野(2004)「日本企業の多角化経営と組織構造」『組織科学』vol.37(3))

		事業数の増減			合計
		減少	維持	増加	
既存主力事業への投資	縮小	6	4	3	13
	維持	26	44	18	88
	拡大	24	26	16	66
合計		56	74	37	167

6/27/07

4

クロス集計の例(2)

■ 独立性の検定

$$C = \sum_{i=1}^k \sum_{j=1}^l \frac{(x_{ij} - e_{ij})^2}{e_{ij}}$$

$$C \xrightarrow{d} \chi^2((k-1) \times (l-1))$$

■ 前表のケース

■ $c=2.98$

■ $\chi^2(4)$ の上位5%有意点は9.4877



帰無仮説 (独立性) を棄却できない

6/27/07

5

相関係数の種類

■ Pearsonの積率相関係数

■ 一般に「相関係数」と呼んでいるもの

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right) \left(\sum_{i=1}^n (y_i - \bar{y})^2\right)}}$$

■ Non-parametricな関係性の指標

■ グッドマン=クラスカルの順序連関係数
Goodman-Kruskal Gamma

■ Spearmanの順位相関係数

■ 順位変数の相関

6/27/07

6

グッドマン=クラスカルの順序連関係数

■ 順序のあるカテゴリー変数の関係性の指標

対となったデータ $(x_1, y_1), (x_2, y_2)$ を比較すると、下表の分類のどこかに入る。

あらゆるデータの対に対して、これを行い

$G = (\#A - \#B) / (\#A + \#B)$ を求める。

	$y_1 < y_2$	$y_1 = y_2$	$y_1 > y_2$
$x_1 < x_2$	A	C	B
$x_1 = x_2$	C	C	C
$x_1 > x_2$	B	C	A

6/27/07

7

#Aの具体的な求め方

■ セル (i, j) を固定し、そのセルより右下にあるセルの度数を N_{ij}^+ とする。

■ n_{11} に対応する N_{11}^+ は $n_{22} + n_{23} + n_{32} + n_{33}$

■ n_{21} に対応する N_{21}^+ は $n_{32} + n_{33}$

	減少	不変	増加
減少	n_{11}	n_{12}	n_{13}
不変	n_{21}	n_{22}	n_{23}
増加	n_{31}	n_{32}	n_{33}

6/27/07

8

#Bの具体的な求め方

- セル(i,j)を固定し、そのセルより左下にあるセルの度数を N_{ij}^- とする。
- n_{12} に対応する N_{12}^- は $n_{21} + n_{31}$
- n_{13} に対応する N_{13}^- は $n_{21} + n_{31} + n_{22} + n_{32}$

	減少	不変	増加
減少	n_{11}	n_{12}	n_{13}
不変	n_{21}	n_{22}	n_{23}
増加	n_{31}	n_{32}	n_{33}

6/27/07

9

クロス表の例

- 標本数 $n=167$ ペアの数 13,861

#A=2556

$$=6 \times (44+18+26+16) + 4 \times (18+16) + 25 \times (26+16) + 44 \times 16$$

#B=2516

$$=4 \times (26+24) + 3 \times (26+24+44+26) + 44 \times 24 + 18 \times (24+26)$$

$$G=(2556-2516)/(2556+2516)=0.0078$$

連関ナシ

6/27/07

10

Spearmanの順位相関係数(1)

- 右の表のデータ

- 標本の大きさ: 5

- Xの平均: 3.2

- Yの平均: 10.8

- Xの2乗和

$$9+16+36+1+4=66$$

- Yの2乗和

$$100+81+225+64+144=614$$

- XとYの積和

$$30+36+90+8+24=188$$

X	Y
3	10
4	9
6	15
1	8
2	12

6/27/07

11

Spearmanの順位相関係数(2)

- まずPearsonの積率相関係数を求める。

$$\sum_{i=1}^5 (X_i - \bar{X})^2 = \sum_{i=1}^5 X_i^2 - 5\bar{X}^2 = 66 - 5 \times (3.2)^2 = 14.8$$

$$\sum_{i=1}^5 (Y_i - \bar{Y})^2 = \sum_{i=1}^5 Y_i^2 - 5\bar{Y}^2 = 614 - 5 \times (10.8)^2 = 30.8$$

$$\sum_{i=1}^5 (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^5 X_i Y_i - 5\bar{X}\bar{Y} = 188 - 5 \times 3.2 \times 10.8 = 15.2$$

$$r = \frac{\sum_{i=1}^5 (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^5 (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^5 (Y_i - \bar{Y})^2}} = \frac{15.2}{\sqrt{14.8} \sqrt{30.8}} = 0.71$$

(Pearsonの)
相関係数

6/27/07

12

Spearmanの順位相関係数(3)

- 下表のデータX, Yを、小さい順に順位をつけ変換し、W, Zとする

X	Y	W	Z
3	10	3	3
4	9	4	2
6	15	5	5
1	8	1	1
2	12	2	4

6/27/07

13

Spearmanの順位相関係数(4)

- 順位データW, Zの相関係数を求める。

$$\sum_{i=1}^5 (W_i - \bar{W})^2 = \sum_{i=1}^5 W_i^2 - 5\bar{W}^2 = 55 - 5 \times 3^2 = 10$$

$$\sum_{i=1}^5 (Z_i - \bar{Z})^2 = \sum_{i=1}^5 Z_i^2 - 5\bar{Z}^2 = 55 - 5 \times 3^2 = 10$$

$$\sum_{i=1}^5 (W_i - \bar{W})(Z_i - \bar{Z}) = \sum_{i=1}^5 W_i Z_i - 5\bar{W}\bar{Z} = 51 - 5 \times 3 \times 3 = 6$$

$$r^* = \frac{\sum_{i=1}^5 (W_i - \bar{W})(Z_i - \bar{Z})}{\sqrt{\sum_{i=1}^5 (W_i - \bar{W})^2} \sqrt{\sum_{i=1}^5 (Z_i - \bar{Z})^2}} = \frac{6}{\sqrt{10} \sqrt{10}} = 0.6$$

Spearmanの
順位相関係数

6/27/07

14

[課題] (本日提出分)
サンプリング 解説

- 一昨年の課題提出例
- 有効答案10
 - 次ページ以降にまとめ
- 母集団分布は右表のとおり

		回答者数	比率
視聴率	視聴率	1,182	39.4%
	携帯		
	不所持	108	3.6%
	D社	1,379	46.0%
	A社	1,149	38.3%
	V社	362	12.1%

6/27/07

15

標本視聴率 (母集団視聴率=39.4%)

視聴率	系統150	層化150	多段150	単純150	単純60
0.25	0	0	0	0	0
0.30	0	0	1	0	2
0.35	3	2	4	0	2
0.40	5	3	2	8	1
0.45	2	5	3	2	3
0.50	0	0	0	0	2

6/27/07

16

