

第7回 統計的分析の基礎

1. クロス表分析の特徴と限界

性別が所得に及ぼす効果についての分析

(1) 性別×所得のクロス表

女性よりも男性のほうが、所得が多くなる傾向がある。

(2) 年齢を統制した場合の性別×所得のクロス表

20歳代を除き、女性よりも男性のほうが所得が多くなる傾向がある。

(3) 学歴を統制した場合の性別×所得

どの学歴階層においても、男性と女性の所得格差は大きい。

では、学歴と年齢を同時に考慮した、四重クロス表を作成することはできないのか？

→原理的には可能だが、サンプル数が足りず、十分な分析ができなくなる。

(クロス表分析は、各セルの期待度数が5以上であることが望ましい)。

また、クロス表の数が多くなりすぎて、全体を要約的につかむことが難しくなる。

クロス表はすべてのデータの分布が見えるので、仮説から逸脱するケースも含めて、全体の傾向をつかむのに適している。

量的データも、カテゴリーに分類することによってクロス表を作成することが可能なので、あらゆるタイプの変数に適用できる。

しかし、4つ以上の変数間の関連を同時に分析することは、サンプル数の限界、表が複雑になりすぎるなどの理由から、難しくなる。この難点を克服するためには量的変数による分析が必要となってくる。

2. 統計的分析の基礎

(1) 代表値

ある統計的集団についての観測値（データ）をひとつの値で代表させるとき、その値を代表値という。代表値には、平均値（算術平均）、中央値（中位数、メディアン）、最頻値（モード）などがある。

平均値（算術平均）は、各個体の観測値の総和を個体数で割ったもの。

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

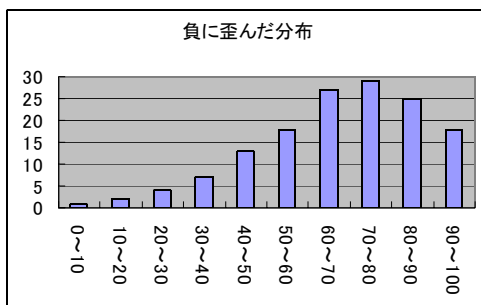
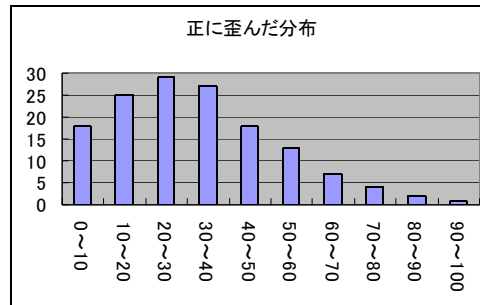
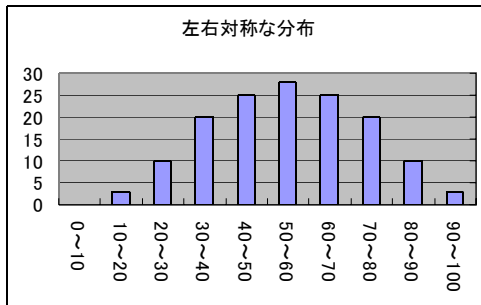
中央値（メディアン）は、各個体を観測値の大きい順に並べたときにちょうど真ん中にくる個体の観測値。

最頻値（モード）は、観測値ごとに個体数を数えたときに、最も個体数の多い観測値。

分布が対称である場合、平均値＝中央値＝最頻値

分布が正に歪んでいる場合、平均値＞中央値＞最頻値

分布が負に歪んでいる場合、平均値＜中央値＜最頻値



架空の数値例；平均値、中央値、最頻値

左右対称な分布； 55 55 55

正に歪んだ分布； 33 30 25

負に歪んだ分布； 68 70 75

(2) 分散

各観測値の平均からの偏差の2乗の平均

$$S_x^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 = \frac{1}{n} \sum x_i^2 - \bar{x}^2$$

分散は、各データの観測値の2乗の平均から、平均の2乗を引いたものに等しい。

$$\begin{aligned} \frac{1}{n} \sum (x_i - \bar{x})^2 &= \frac{1}{n} \sum (x_i^2 - 2\bar{x}x_i + \bar{x}^2) \\ &= \frac{1}{n} \sum x_i^2 - \frac{1}{n} \cdot 2 \cdot \bar{x} \sum x_i + \frac{1}{n} \sum \bar{x}^2 \\ &= \frac{1}{n} \sum x_i^2 - 2\bar{x}^2 + \bar{x}^2 \\ &= \frac{1}{n} \sum x_i^2 - \bar{x}^2 \end{aligned}$$

(3) 標準偏差

分散の平方根

$$S_x = \sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2}$$

(4) 標準測度

各個体の観測値から平均値を引いたものを平均値からの偏差という。標準測度は平均値からの偏差を標準偏差で割ったもの。

平均からの偏差を d_i とすると、 $d_i = x_i - \bar{x}$

平均からの偏差の平均は 0 $\bar{d} = 0$

標準測度を z_i とすると、

$$z_i = \frac{d_i}{S_x} = \frac{x_i - \bar{x}}{S_x}$$

標準測度の平均は 0、分散は 1。

$$\bar{z} = \frac{1}{n} \sum \frac{x_i - \bar{x}}{S_x} = \frac{\bar{d}}{S_x} = 0$$

$$S_z^2 = \frac{1}{n} \sum (z_i - \bar{z})^2 = \frac{1}{n} \sum \left(\frac{x_i - \bar{x}}{S_x} \right)^2 = \frac{1}{n} \sum \frac{(x_i - \bar{x})^2}{S_x^2} = \frac{S_x^2}{S_x^2} = 1$$

次回までの宿題。

<http://www.rikkyo.ne.jp/web/yamatsumoto/index.html> → Office → 社会学データ実習に掲載されている『Σの世界』をダウンロードして、Σの用法について復習しておくこと。