第11回　平均値の比較と相関比

独立変数 x が質的変数で、従属変数 y が量的変数の場合、x のカテゴリーごとに y の平均値を比較して、違いがあるかどうかを検討することができる。

x のカテゴリーごとの y の平均値を $\overline{y_i}$ と表す。i はカテゴリー番号。

質的変数と量的変数との関連を見るのに、**相関比**が用いられる。

相関比 η^2 の定義（まず、平方を考える）。

$$\eta_{yx}{}^2 = \frac{x\text{を与えたときの}y\text{の平均値の重みつき分散}}{y\text{の全分散}} = \frac{S_{\overline{yi}}{}^2}{S_y{}^2}$$

$$= \frac{\dfrac{1}{n}\displaystyle\sum_{i=1}^{k} n_i(\overline{y_i}-\overline{y})^2}{\dfrac{1}{n}\displaystyle\sum_{i=1}^{k}\sum_{j=1}^{ni}(y_{ij}-\overline{y})^2} \qquad \cdots ①$$

ところで、$\eta^2{}_{yx}$ は　$1-\dfrac{x\text{を与えたときの}y\text{の分散}}{y\text{の全分散}}$　に等しい。

$$1-\frac{\dfrac{1}{n}\displaystyle\sum_{i}^{k}\sum_{j}^{ni}(y_{ij}-\overline{y_i})^2}{\dfrac{1}{n}\displaystyle\sum_{i}\sum_{j}(y_{ij}-\overline{y})^2}$$

$$= \frac{\dfrac{1}{n}\sum\sum(y_{ij}-\overline{y})^2-\dfrac{1}{n}\sum\sum(y_{ij}-\overline{y_i})^2}{\dfrac{1}{n}\displaystyle\sum_{i}\sum_{j}(y_{ij}-\overline{y})^2}$$

分母を $S_y{}^2$ と表記し、ΣΣまでを括り出す。

$$= \frac{1}{S_y{}^2}\cdot\frac{1}{n}\sum_{i}\sum_{j}\{(y_{ij}-\overline{y})^2-(y_{ij}-\overline{y_i})^2\}$$

｛ ｝内を展開する。

$$=\frac{1}{S_y^2}\cdot\frac{1}{n}\sum_i\sum_j\left(y_{ij}^2-2y_{ij}\bar{y}+\bar{y}^2-y_{ij}^2+2y_{ij}\bar{y}_i-\bar{y}_i^2\right)$$

( )内の項を整理して

$$=\frac{1}{S_y^2}\cdot\frac{1}{n}\sum_i\sum_j^{n_i}\left(\bar{y}^2-2y_{ij}\bar{y}+2y_{ij}\bar{y}_i-\bar{y}_i^2\right)$$

ここで、

$\sum_j^{n_i} y_{ij}=n_i\bar{y}_i$　　（級内の総和）　であることを利用して、Σ j をカッコ内に入れる。

$$=\frac{1}{S_y^2}\cdot\frac{1}{n}\sum_i\left(n_i\bar{y}^2-2n_i\bar{y}_i\bar{y}+2n_i\bar{y}_i\bar{y}_i-n_i\bar{y}_i^2\right)$$

$$=\frac{1}{S_y^2}\cdot\frac{1}{n}\sum_i n_i\left(\bar{y}^2-2\bar{y}_i\bar{y}+\bar{y}_i^2\right)$$

$$=\frac{\frac{1}{n}\sum_i n_i(\bar{y}_i-\bar{y})^2}{\frac{1}{n}\sum\sum(y_{ij}-\bar{y})^2}$$

つまり、

$$\frac{x\text{を与えたときの}y\text{の平均値の重みつき分散}}{y\text{の全分散}}=1-\frac{x\text{を与えたときの}y\text{の分散}}{y\text{の全分散}}$$

よって、つぎのような関係が成り立つ。

$$\frac{\frac{1}{n}\sum_i n_i(\bar{y}_i-\bar{y})^2}{\frac{1}{n}\sum_i\sum_j(y_{ij}-\bar{y})^2}+\frac{\frac{1}{n}\sum_i\sum_j(y_{ij}-\bar{y}_i)^2}{\frac{1}{n}\sum_i\sum_j(y_{ij}-\bar{y})^2}=1$$

分母を払うと、

$$\frac{1}{n}\sum_i n_i(\bar{y}_i-\bar{y})^2+\frac{1}{n}\sum_i\sum_j(y_{ij}-\bar{y}_i)^2=\frac{1}{n}\sum_i\sum_j(y_{ij}-\bar{y})^2$$

つまり、級間分散＋級内分散＝全分散

n 倍した式では、級間変動＋級内変動＝全変動

$$SSB + SSW = SST$$

結局、相関比とは、

$$\sqrt{\frac{SSB}{SST}}=\sqrt{1-\frac{SSW}{SST}}$$

・この平方が、全分散にたいする級間分散の説明率を示す（決定係数)。
・級内変動 SSW が小さければ、相関比は 1 に近づき、大きければ 0 に近づく。

【例題】学歴 q43 と仕事による収入 income2 の平均値および学歴と収入の相関比を計算する。(学歴は質的変数、収入は量的変数であることに注意)。

spss のコマンドは、

```
means tables income2 by q43 /statistics anova.
```

報告書

income2

q43 学歴	平均値	度数	標準偏差
1 中学卒	277.8689	61	221.26993
2 高校卒	354.6099	282	307.37673
3 短大卒	306.2992	127	249.92000
4 大学卒	576.5781	301	402.21211
合計	427.2374	771	355.49127

分散分析表

			平方和	自由度	平均平方	F値	有意確率
income2 x q43 学歴	ｸﾞﾙｰﾌﾟ間	(結合)	11419049.476	3	3806349.825	33.991	.000
	ｸﾞﾙｰﾌﾟ内		85888966.088	767	111980.399		
	合計		97308015.564	770			

連関の測定方法

	ｲｰﾀ	ｲｰﾀの2乗
income2 x q43 学歴	.343	.117