

現代日本学演習 II 「統計分析の基礎」

第8講 平均値の比較

田中重人 (東北大学文学部准教授)

[テーマ] ふたつのグループ間での平均値の比較

1 度数分布表のオプション

度数分布表の「統計」オプションで「平均」と「標準偏差」をチェック。

- 「記述統計量」 「記述統計量」でも出力できる。(そのほかにもさまざまな分析の「記述統計」オプションで出力できる)
- PSPP などの統計ソフトの「標準偏差」は、統計学の普通の定義とはちがう式で計算している(教科書 p. 48 注6)。データが大きくなれば(およそ 200 以上なら)このことによるちがいはほとんどなくなるが、小さいデータ(たとえば 10 人程度)では大きなちがいになるので注意。

練習問題: 前回宿題について、PSPP にデータを打ち込み、平均値と標準偏差を出力してみよう。

2 順序尺度の変数の「平均値」

平均値は、本来は、間隔尺度以上の水準の変数にしか使えない。しかし、実際には、一定条件を満たせば、順序尺度についても平均値をとっていいとする基準が使われている。

- 潜在的には間隔尺度のはず
- 測定のポイントが一定間隔

具体的には、4 点以上の尺度であって、正規分布に近似している場合(教科書 p. 53-59)。これは、「偶然の積み重ねで形成されるものは正規分布にしたがう」という仮定による。

「正規分布に近似」しているかどうかは、通常、つぎの 3 点で判断する。

- 単峰性
- 左右対称性(歪度)
- 中央への集中度(尖度)

PSPP 度数分布表(または棒グラフ)で検討するとよい。

「度数分布表」の「統計量」オプションで「歪度」「尖度」を指定すると、正規分布との乖離度を統計的に検討できる。これらの値は、正規分布のとき 0 をとり、絶対値が大きくなるほど、正規分布から外れる。およそ ± 2 の範囲を超えていれば、正規分布からのずれが無視できない。

練習問題: 配布しているデータ中のさまざまな意識変数について、度数分布表(棒グラフ、歪度、尖度)を出力し、正規分布に近似しているかどうかを判断してみよう。

これらの条件を満たさない場合は非線形変換 (教科書 p.142-144) をおこなったり、順位に変換したりすることがある。あるいは、平均値を使わずに中央値を使って分析することもある。

なお、2 値の変数は、この条件にかかわらず間隔尺度とみなしてよいが、一定以上のデータ数があり、あまり偏っていないことが必要。

3 平均値の欠点

平均値は「はずれ値」(outlier) の影響を受けやすい。あまりにかけはなれたケースがあるときは

- 上下数%を取りのぞく (調整平均: 教科書 p. 46)
- 順位に変換したり中央値を使って分析

などの方法を使うことがある。

また、極端なはずれ値がなくとも、左右非対称の分布の変数 (所得、人口、めったに起こらない現象の経験回数など) では、平均値より中央値の方が適切な代表値であることが多い。

4 ふたつのグループ間での平均値の比較

データをグループに分けて、それぞれ平均値 (= 層別平均) を求め、それらの間の差をもとめる。この差の大きさを、標準偏差を基準にして評価する。具体的には、effect size (ES) または 相関比 (η^2 : イータ) という統計量を使う。

4.1 PSPP でのグループ別平均値の出力

メニューの「分析」から「平均の比較」 「グループの平均」を開く。

- 「従属変数」に平均値を計算する変数を指定
- 「独立変数」にグループの変数を指定

4.2 エフェクト・サイズ

Effect size (ES) : 一般には「Cohen の d 」と呼ばれる。

$$ES = \frac{\text{グループ別平均の差}}{\text{併合 SD}} \quad (1)$$

「併合 SD」の計算については教科書 p. 137 [式 5-8] の分母を参照。大雑把には、グループ別の SD の中間の値と考えてよい。

ES は、計算が簡単であり、直感的に把握しやすい。しかし、各グループの人数を考慮せず平均値だけ比較するため、グループの人数が大きくちがう場合でも、同じ人数に 2 等分されている場合でも、その間のちがいは ES の値に反映しない。また、2 グループ間の比較だけを行うものであるため、3 つ以上のグループを比較するのにはつかえない。

ES は PSPP では計算できない。

4.3 相関比 (correlation ratio)

- 各グループの個体が全員そのグループの平均値を持つ状況を仮定してSDを求める。
- この仮想SDを実際のSDで割った数値が「相関比」である。数式では η (eta) であらわす

イータ (η) は 0~1 の範囲の値をとり、独立変数の影響力をあらわす

PSPPでは、クロス表の「統計」オプションで「イータ」を選択する。出力には2種類の値が出てくるが、平均値を計算したい変数が「従属」になっているほうを見る。

5 課題

- (1) 適当な変数について、度数分布表・平均・標準偏差を出力（全体と男女別）
- (2) 性別でわけて度数分布を折れ線グラフで描く
- (3) これらの分析結果から何が言えるか、解釈を書く。