

日本赤十字九州国際看護大学/Japanese Red

Cross Kyushu International College of

Nursing

分散分析とF検定 ; 2

メタデータ	言語: Japanese 出版者: 公開日: 2020-08-04 キーワード (Ja): キーワード (En): One-way ANOVA, Excel calculation, js-STAR calculation 作成者: 守山, 正樹 メールアドレス: 所属:
URL	https://jrckicn.repo.nii.ac.jp/records/725

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 International License.



ここで平均に明らかな差があるということも大切ですが、それよりも注目すべきは、四つのグループがそれぞれの平均と狭い範囲でのバラツキ・変動を示し、明らかな特徴を持つグループとして、際立っていることです。学習意欲という元データの全変動（SSTotal）が、主として、動画の時間という要因による変動（SSTreatment）で説明できることが、読み取れます。特に短い動画で学習意欲が高いということであれば今後より短時間に集中して学生の皆さんに多くのことを考えてもらうような形の動画を作成することが大切であるとわかります。

3 計算演習

前項の動画の長さを例にした実験計画の例です。

15名の学生をランダムに3群に分け、各群に異なる長さの動画を視聴してもらい、視聴後に学習意欲（1～100点）を測定したところ、表の結果を得た。動画の長さが学習意欲に影響を与えているかを、平均に差があるかを、有意水準0.05で検定しなさい。

動画 10 分	動画 30 分	動画 90 分
80	60	50
85	75	55
70	52	70
78	58	40
60	68	45

1) エクセルを用いる場合

- すでに設定したエクセルの分析ツールを使うためには、まずエクセルの画面上部にあるメニューからデータのタブを選びます。
- すると、上の右端に分析というタブが現れるので、それをクリックします。
- すると、分析ツールのボックスが現れるので、メニューから「分散分析：一元配置」を選びOKを押します。
- すると「分散分析：一元配置」のボックスが現れるのでまず「入力範囲」を指定します。
- 各群のデータは縦方向に並んでいますので、「列」を選びます。
- 入力範囲の先頭行はデータではなく、「動画 90 分」など群の名前ですので、「先頭行をラベルとして使用」もチェックします。
- 指定してOKを押すと結果が二つの表で示されます。
- 概要の表には、三つあった群別の標本数、合計、平均、分散が示されます。
- 次の分散分析表で、まず変動とあるのは、偏差二乗和（ ss ）です。データ全体のSSが。グループ間（群間）のSSとグループ内（群内）のSSに分けて示されます。
- 表の自由度の部分を見ると「2、12、14」と数値が三つ並んでいます。14は「全体の自由度」、「全標本数-1」で計算されます。他方、2は要因の群分けに関する自由度で、「要因の自由度」とよび「グループ数-1」で計算されます。「全体の自由度」から「要因の自由度」を引き算したものが「残差の自由度」です。
- F分布の形は二つの自由度で規定され、印刷されたF分布表を見るときも、二つの自由度を用意する必要があります。

- ・F 値とは「グループ間の分散」を「グループ内の分散」で割り算したのが「観測された分散比」、これが F 値です。
- ・F 値 6.20446 に対応する有意確率 (P 値) は 0.014119 です。よって、今回の計算例では p 値 (有意確率) は、有意水準を 5% (0.05) と設定した場合には、それより低い値をとるため、帰無仮説は棄却されます。

2) js-STARによる場合

- ・js-STAR のサイトを開くと左側にメニューが出てきます。
- ・js-STAR には分散分析として多くの選択肢が用意されていますが、その最初、As (1 要因参加者間) を選びます。
- ・まず現れるのは、群が二つ、各群の参加者二名だけの単純な表です。この表は条件に応じて縦方向に拡張して使うようになっています。2 と出ている部分に、実際の比較したい群の数、各群の参加者人数を入力します。先ほどの例で言えば第 1 群から第 3 群まで、各群の参加者は 5 名です。
- ・条件を入力すると、表が縦に伸びて値が入力できるようになります。ここに比較したい各群のデータ (学習意欲) を群別に入力します。
- ・入力を終えて「計算のキー」を押すと結果が現れます。
- ・js-STAR の計算結果は、表示窓には計算結果の一部分しか出ていませんので、カーソルを動かして最初の部分から結果を見ていきます。
- ・先ほどのエクセルでの計算とほぼ一致する結果が得られたことを確認してください。
- ・エクセルでは「観測された分散比」となっていた項目は、F (F 値) として示されています。
- ・F 値 6.20 の横に * 印がついています。これはその下の説明によると「* $P < .05$ 」つまりこの F 値 (6.20) に対応する有意確率 (p 値) は 0.05 (有意確率) より小さいこと、よって有意水準 0.05 で帰無仮説は棄却されることを意味します。

4 分散分析の背景

1) 分散分析の歴史

分散分析が形になったのは 1918 年にロナルド・フィッシャーが分散という用語を導入してからです。その後分散分析はフィッシャーが 1925 年に書いた本を介して、広く世界に知られるようになりました。しかし分散分析に至る考え方は何世紀にもわたって育まれてきたとされ、統計学に関する様々な考え方「仮説検定、二乗和の分割、実験手法、加法モデル」などが含まれています。

特に「全体の変動 (偏差二乗和、SS) をグループ内 SS とグループ間 SS に分解する」、「各 SS を自由度で割り算して分散を求める」、「二つの分散の比を F で表わす」という論理は、計算方法がシンプルで分かりやすく、多くの場面で用いられました。

2) F 分布の歴史

F 分布は分散分析を行うときの統計量として知られています。F 分布においては、この考え方への貢献が大きい 2 人の統計学者、スネディガーとフィッシャーの名前をつけて、F-distribution、Snedecor's F distribution、the Fisher-Snedecor distribution などと呼ばれています。

「全体の変動（偏差二乗和、SS）をグループ内SSとグループ間SSに分解する」を数式で書くと以下ようになります。

$$SS_{total} = SS_{error} + SS_{treatments}$$

なお、元データの全変動中、要因による変動を除いた残りの変動がSS_{error}です。

また分散分析で使うF値とは、要因による変動の分散を、残りの変動の分散で割り算したものです。

3) 質的研究と分散分析、発想の違い

看護研究では研究方法として質的な研究の方がよく使われる傾向にあります。特にこの分散分析の方法は、質的な研究の研究者にとっても興味深い研究方法だと考えられます。研究方法としての特徴は、名前にも現れている通り、分散に注目して研究する方法です。

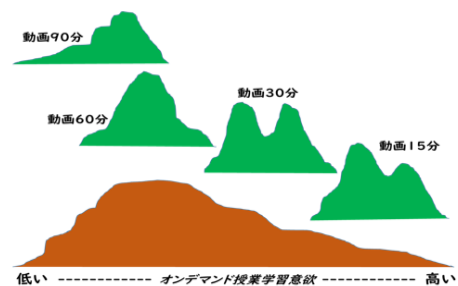
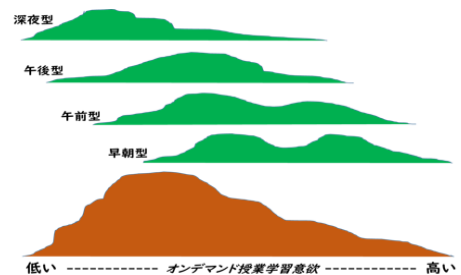
質的な研究の場合は少数の人々の思考や行為における意味に注目します。他方、分散分析では、授業第2回目で平均値や標準偏差などと共に説明した基本的な統計量「分散」に注目します。質的な研究が個人的な現象の意味から入ることが多いのに対し、分散分析では意味よりも、社会集団における現象の分布の形に注目します。

先ほどの分布図を思い出してください。

F値や有意確率は計算しないと求められません。

しかし図を見れば、どちらの群分けの方が現象を解明する手がかりが得られるか、明白ですよね。生活時間帯で分けた上の図では分布の重なりが多く、全体の変動から各要因の変動を明らかに取り出すのは困難です。

他方、下の図では、動画の長さで群分けすることで、各動画の変動がはっきりと分離できます。この場合にF値を計算したら、おそらく、確実に帰無仮説は破棄され、4群の間に差があると結論できるでしょう。



5 まとめ

さて今日は分散分析についてお話ししました。

分散分析は統計学の考え方が最もよく現れた分析方法です。みなさんも、ぜひ試してみてください。

演習問題

1. 分散分析は統計学の代表的な分析方法です。あなたはどのような課題に分散分析を使ってみたいですか。50字以内で書いてください。
2. 動画で分散分析の計算方法を学んでください。エクセルを用い、my 標本で何か分散分析を行い、結果を50文字以内で書いてください。(my 標本ではなく150名データを用いても構いません) エクセルが利用できない場合は、動画中のエクセルの説明を見て、感じたことを50字以内で書いてください。(今エクセルを使えなくても、登校禁止が解除されたら、ぜひ情報処理室でエクセルに触れてください。)
3. js-STAR で分散分析を行ってください。

<http://www.kisnet.or.jp/nappa/software/star/>

すでにエクセルで計算済みであっても、js-STARでも計算を試みてください。同じデータを用い、複数の方法で計算してみることで、各方法の特徴を把握でき、また各方法の限界も理解できます。結果や気づいた点を50字以内で書いてください。