

日本赤十字九州国際看護大学/Japanese Red Cross Kyushu International College of Nursing

Analysis of variance and F-test ; 1

メタデータ	言語: jpn 出版者: 公開日: 2020-08-04 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	https://jrckicn.repo.nii.ac.jp/records/724

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



第 12 回 分散分析と F 検定

皆さんこんにちは。今回は分散分析についてお話しします。

1 分散分析の考え方

1) データのばらつき・変動から出発

分散分析は「偏差二乗和 (SS) と分散 (VAR 、 σ^2)」が大活躍する分析方法です。偏差二乗和や分散って何か覚えていますか。データの「ばらつき・変動」を示す基本的な値です。第 3 回目の授業でお話しました。実測値と平均の差、偏差もデータのばらつきを表わすのですが、集団全体について偏差を合計するとゼロになってしまいます。そこで、ゼロにならないように二乗して合計したのが偏差二乗和 (SS)、その平均が分散 (σ^2) です。ワークシートで計算したことを思い出してください。

ただ一つ重要な相違点があります。それは、今回の分散分析で用いる分散は、実は第 3 回目の授業での分散とはやや異なり、偏差二乗和を割り算するとき、データ数 N ではなく自由度 $n-1$ を用いるという点です。

2) 分散分析の種類

検討する要因が一つだけの場合が一元配置分散分析です。同時に検討する要因の数が増えると、二元配置分散分析、多元配置分散分析などになります。

以下では、もっとも一般的な「一元配置分散分析」を取り上げます。

3) 何に使うか

分散分析は統計学を使って本格的に研究をするときに役立つ方法です。探索的な使い方と実験計画的な使い方とがあります。

- ・探索的な使い方；何らかの観測データが得られた場合、そのデータのばらつき・変動は、どのような要因によって引き起こされているか、可能性のある要因を探索するような使い方です。例えば新型コロナウイルスによる発症の重症度というデータがあったとします。それはどのような要因で影響を受けるのか、年齢・性別・栄養状態・民族的などの要因から、一つまたは二つを選び、要因でグループ分けし、グループ間やグループ内で、要因の効果をデータの平均や分散を用いて比較します。
- ・実験計画的な使い方；あらかじめいくつかの処理・条件を設定した実験計画を立て、実行したときに、処理群間や処理群内で効果を、平均や分散を用いて比較します。たとえば、新たな肥料 A、B、C について、また新たな治療薬 A、B、C の効果を比較するなどの使い方です。
- ・コンピューターが現在ほど進歩するまでは、分散分析は基本の方法でした。その後コンピューターが進歩し計算が容易になるにつれて、分散分析は、他のより大量な計算を必要とする方法と組み合わせて使われることが増えています。その結果、最近の統計学の教科書では分散分析という項目が登場しない場合もあります。

4) なぜ分散に注目するか

なぜデータのばらつき・変動を示す偏差二乗和や分散が、分析や研究の出発点になるのでしょうか。それはデータ全体を何かの要因でグループ分けしてグループ別（群別）の分散を求めた時、その要因によるグループ分けが意味を持っている程度に合わせて、「データ全体の分散」に対し「要因によって説明できる分散」の占める割合が、変化するからです。他方、何らかの要因でグループ分けをしたとしても、そのグループ分けによって説明できる分散の部分が少なければ、グループ分けしたことに意味がない、ということになります。

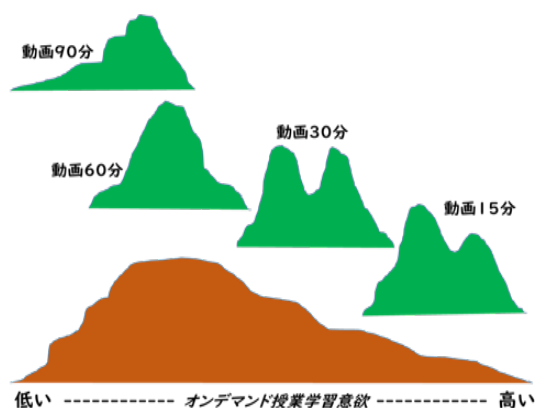
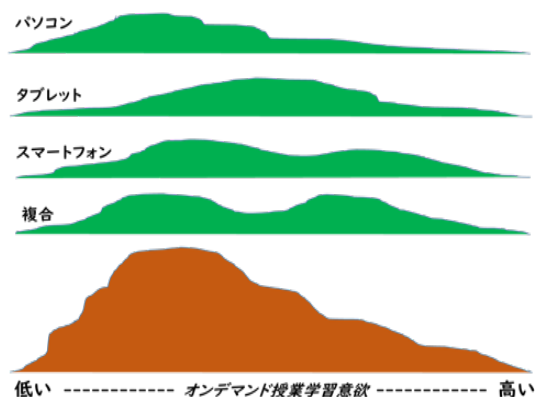
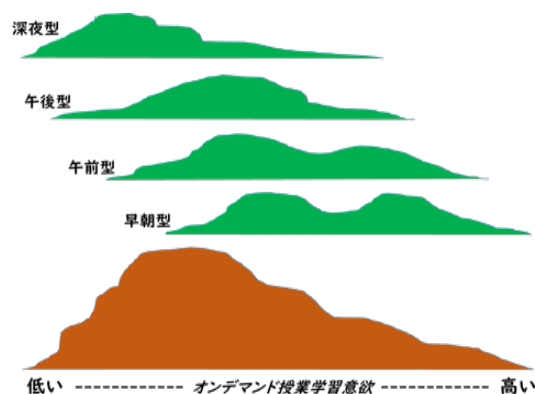
2 グラフで考える

分散分析の意味を直感的に把握できるよう、グラフで説明します。「COVID-19 禍のもとでのオンデマンド授業についての学生の皆さんの学習意欲」を出発点となるデータ全体として考えてください。まず、学習意欲を0から100までの数値として、その分布を茶色のヒストグラムで表わします。学習意欲は低い人から高い人まで様々です。この学習意欲はどのような要因によって説明されるのでしょうか。いくつかの要因でグループ分けして考えてみます。

図1では、学生の皆さんがオンデマンド学習を行う主な時間帯で、早朝型・午前型・午後型・深夜型とグループ分けしました。時間帯が早い方が遅い方に比べて、学習意欲の平均が高い傾向が認められます。他方、どの型でもデータのばらつき・変動の幅が広めで、4つの分布は相当に重なっています。

図2では、オンデマンドの動画を見ている方法でグループ分けしました。パソコンよりもタブレット端末が多少学習意欲の平均が高いようです。スマートフォンを使う人と、幾つかの方法を複合して使う人とは、差ははっきりしません。他方、どの方法でもばらつき・変動の幅が広く、全域にわたって重なっています。

図3では、動画の時間の長さでグループ分けしました。ここで大きな差が出てきました。長い動画を見ている人は、動画が途中で途切れたり長すぎて注意が散漫になるのか学習意欲が低くなっています。他方、動画の時間が15分と極めて短いと、見る時の集中力が高まるせいか、学習意欲の平均が高くなっています。



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（S S T o t a l）が、主として、動画の時間という要因による変動（SS Treatment）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（s s）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。