



2021年4月

F分布と等分散性の検定【第54回生物統計学】

1 F分布と等分散性の検定

推測統計学においては、標本を元にして母集団の平均や分散を推測するため、標本の平均や分散を確率分布として捉える考え方が必要になってきます。正規分布に従う二つの独立な母集団の分散が等しいとき、それらの集団の標本の分散の比が従う確率分布が「F分布」です。それぞれの集団における標本の偏差平方和が、自由度 m のカイ二乗分布に従う確率変数 Y 及び自由度 n のカイ二乗分布に従う確率変数 Z を用いて、

$$X = \frac{(Y/m)}{(Z/n)} \dots (1)$$

で表される確率分布 X を、「自由度 (m,n) の F 分布」と呼びます。分母分子ともにカイ二乗分布が含まれているため、分散の比をとっていることが分かります。二つの母集団の分散の比を比べているため自由度を二つ用いて説明しなくてはならないのが特徴です。

母分散 σ_A の正規母集団 A から m 個のサンプルを抽出して算出した不偏分散が U_A^2 、母分散 σ_B の正規母集団 B から n 個のサンプルを抽出して算出した不偏分散が U_B^2 の場合、

$$\frac{(m-1)U_A^2}{\sigma_A^2}, \frac{(n-1)U_B^2}{\sigma_B^2} \text{ はそれぞれ自由度 } m-1, n-1 \text{ のカイ二乗分布に従い、}$$

$$\text{これを(1)式に当てはめると } \frac{\frac{(m-1)U_A^2}{\sigma_A^2}}{\frac{(n-1)U_B^2}{\sigma_B^2}} \text{ が自由度 } (m-1, n-1) \text{ の F 分布に従います}$$

上記の通り、正規分布に従う集団においては(偏差平方和)/(母分散)が自由度(サンプル数-1)の正規分布に従うため、二つの母集団の分散が等しい ($\sigma_A = \sigma_B$) と仮定すると、両者の不偏分散の比が自由度 $(m-1, n-1)$ の F 分布に従います。「二つの母集団の分散は等しい」という帰無仮説に基づいて二つの母集団の等分散性を検定する手法が F 検定です。

二つの集団の母平均の差の検定に用いる最もオーソドックスな手法である Student の t 検定は、両者が等分散の正規分布に従うことを前提としており、F 検定で有意差がつかない場合には Welch の t 検定や Mann-Whitney の U 検定を用いたほうが適切です。

検定に際しては F 分布表を用いることが多いです。分母と分子の自由度が交錯する点が有意水準の境界になっているため、不偏分散の比が上側の境界を上回る、あるいは下側の境界を下回るとき有意差が検出でき、母分散が異なっていると判断できます。分布表に記載の無い自由度の組み合わせでも、エクセルの F.INV 関数“=F.INV(有意水準(下側は α 、上側は $1-\alpha$), 分子の自由度, 分母の自由度)”で容易に算出できます。



2 総括

統計解析にあたっては、対象となる集団がどのような分布を取っているのか理解して最適な手法を選択することが求められます。一般に「対称的なふるまいをするもの、先天的な要素が強いもの」について正規性が仮定できる場合が多いとされていますが、常に正規性、等分散性を量的に示した上で解析計画を立てることを意識しましょう。