

P値

P値の意味

今、異なる薬品を投与した2つの動物サンプルからデータを収集したとします。個々の動物の血漿中の酵素を計測した結果、平均値は異なっていました。それが投与した薬品によるものかどうか、すなわち2つの集団は異なる平均値を持っているのかが問題になります。

異なるサンプル平均が観測されたというだけでは、2つの集団の平均値が異なっていたと結論付ける上での説得力に欠けます。2つの集団は同じ平均値を有していた(すなわち薬品は測定対象の酵素に何の影響も与えなかった)が、観測されたサンプル平均の違いはたまたまの偶然で起こったということもあり得ます。観測された違いが真の違いによるものか、単なるサンプリングの過程で生じた偶然か、確認する手立てはありません。やれることと言えば確率を計算することだけです。

最初のステップは帰無仮説、すなわち薬品は計測に何の影響も与えなかったという仮説を立てることです。

P値とはこの恐らくは考えたこともない質問に答える確率のことを言います。

このサイズの実験において、仮に母集団が同一の平均値を有していたとして、少なくとも観測された程度のサンプル平均の差を生ずる確率はどれだけか？

P値に対する誤解

多くの人はP値の意味を誤解しています。今、2つの平均値を比較し、0.03というP値を得たしましょう。

P値の正しい定義は次のとおりです。

2つの集団の平均が同一であったとしても(帰無仮説が真だとしても)、観測された程度の差を生じる確率は3%ほどある。

または

同一の集団からランダムなサンプリングを行った場合、97%の実験では観測されたものより小さい差を生ずるであろうし、3%の実験においては観測されたものより大きな差が生じるであろう。

誤った解釈

観測された差が集団間の真の違いを反映している確率は 97%であり、差が偶然によるものである確率は 3%である。

この最後のステートメントは良く見られる誤りです。正しい定義との違いが鮮明でないという場合には Bayesian perspective をお読みください。

片側 P 値、両側 P 値

2つのグループを比較する場合、片側 P 値 (one-tail/one-sided P values) と両側 P 値 (two-tail/two-sided P values) の違いに注意する必要があります。

片側 P 値にしる両側 P 値にしる同一の帰無仮説 - 2つの集団には差がなく、サンプル平均で観測された差は偶然によるものである - に基づいています。

Note: この例は2つのグループの平均値を比較する対応のない t 検定 (unpaired t test) に対するものです。同様の考え方は他の統計検定にも適用できます。

両側 P 値

両側 P 値は次の設問に答えるものです。

帰無仮説が真であるとして、ランダムに選択されたサンプルが観測されただけ(あるいはそれ以上の)開きのある平均値を持つ確率はどの程度か。

片側 P 値

片側 P 値を解釈するためには、データ収集に先立ちどちらのグループが大きな平均値を持つかが言えなくてはなりません。片側 P 値は次の設問に答えるものです。

帰無仮説が真であるとして、ランダムに選択されたサンプルのうち指定された側が観測されただけ(あるいはそれ以上の)開きのある平均値を持つ確率はどの程度か。

過去のデータなり物理的制約なり、あるいは常識なりによって、差がどちらの方向に生じるかが言える場合にのみ片側 P 値は有効です。ここで問題となるのは差の存在を期待できるかどうかではありません。増加なり減少なりを同じように解釈できるかどうかです。

片側 P 値は次の2つの条件が共に成立する場合のみ選択してください。

どちらのグループが大きい平均値なり比率なりを持つかがデータ収集以前に予測できること

仮に想定外の方がより大きな平均値を持った場合でも、それは偶然によってもたらされたものであり、その差は“統計的には有為でない“と言えること