

効果量

効果量、サンプルサイズ、有意水準、検定力

検定をして有意：そうなる確率は低い：そうじゃないんじゃないか、と考える

そうなる：「帰無仮説」：結局捨てる仮説：言いたいことと逆の仮説

有意になる確率：

サンプル数が多くなればなるほど、p 値は、小さくなる。

有意になる確率が高くなる

p 値は、サンプルサイズに反比例

「平均の差を」標準化した指標：効果量

標準偏差と平均値

平均値から、標準偏差いくつ分離れているか

効果量 1 であれば、1SD 分だけ離れている。

d=0.3 効果量小さい = 実質的な差は小さい

理想は、実質的な差を示す効果量大、かつ、統計的に有意差あり

有意水準 0.05

ふつう起こりえない確率

見方を変えると、

ふつう起こりえない確率：でも絶対起きないわけではなく、その確率の分だけはおきてしまうことがありうる。

起きないはず、なのに、起きることがある。

起きないと思っているのに、それが起きてしまう確率。

それが起きてしまう：間違えてしまう。

何を間違えるか

差がある、と結論付けてしまうのだが、

実は、差がなかった、という間違い

: Type I error

逆の間違いも考えられる

実は差があるのに、差はないと間違えてしまう

: Type II error

この間違いの水準は、0.2 (20%) に設定されることが多い。

検定力：間違いなく判断できる確率

1- α に設定されることが多い (Cohen 1988 が推奨)
は通常、0.2、ということは、検定力は 0.8

検定力が 0.8 というのは、 β が 0.2 ということ、つまり、
「差があるのに、差はないと間違えてしまう」確率が 20% ある。

通常、検定では以下の三つが決まっている

有意水準 0.05

0.2

検定力 0.8

あとは、サンプルサイズと効果量を考える必要がある。

サンプルサイズが小さいと検定力が下がる

サンプルサイズが大きいと検定力が上がる

検定力が下がると、 β が大きくなる。

= 差があるのに、ないと思ってしまうがちになる

検定力が上がると、 β が小さくなる。

= 差が無いのに、あると思ってしまうがちになる

検定力は上がればよいわけではないということが混乱の原因

豊田 2009:35

望ましいのは、サンプルサイズは小さく、かつ、検定力を大きく。

事前に、有意水準、検定力、効果量をきめて、サンプルサイズを求める。

事後で、有意水準、効果量、サンプルサイズから、検定力を調べる。

pwr

<https://cran.r-project.org/web/packages/pwr/vignettes/pwr-vignette.html>

power.t.test(n=

判断目安

<https://kyoto-edu.sakura.ne.jp/?&course=statistics&content=effectSize>

サンプルサイズの決め方

<https://psych.or.jp/publication/world085/pw15/>