Analyzing Linguistic Data 2008/11/13 野村真理子

4. Basic statistical methods

＊*test statistic*(検定統計量)：分布が分かるもの

null-hypothesis（帰無仮説）の検定：関数 **pt( ), pf( ), pchisq( )**

p 値により検定統計量が極端な値かどうかを示す

＊一般的には、確率が 0.05 より低くなると極端な値だとみなすと仮定されているが、有意性の判定に関しては意見は異なる。

＊α *level*（α水準）：*significance level*（有意水準）

検定統計量を極端な値だとみなす cutoff probability

R：α水準 0.05→＊ 0.01→＊＊ 0.001→＊＊＊

p 値があらかじめ設定したαより小さいなら、「統計的に有意である」と言われる。

αを 0.05（ほとんどの言語学、心理言語学のジャーナルで使用）に設定したら、*p*<0.05 か p>0.05 かを報告すればよい。

＊ options (show.signif.stars=FALSE) →有意性の印を無効にする

0.05 のような cutoff point は全く恣意的である。

多くの研究者は、p 値を驚きの測定だと解釈する方を好み、*p*<0.10, *p*<0.05 と報告する代わりに、*p*=0.052, *p*=0.048 と報告する。→これがどれだけ驚くべきことなのかを自分で決めることができる。

＊5%水準で有意である結果を示す論文が出版を認められるとき、それは将来の研究で再現される機会を持つ新しい理論的可能性を開いているからにすぎない。

p 値が小さいほど、実験の power が大きいほど（被験者、項目、反復などの数が大きいほど）、複製研究が結果を証明する可能性はより大きくなる。

＊社会科学においては、*p* 値 0.05 が統計的に有意であるとみなすのはふさわしい。

さまざまな要因を完全にコントロールすることがむずかしい。

cf. 物理学→*p* 値を非常に小さくすることが可能

＊*one-tailed test*（片側検定）：方向性のある検定

（例）語の使用頻度の効果についての一連の研究：高頻度語は低頻度語よりも早く認識される傾向がある。

→頻度が予測変数である実験をすると、高頻度語に対しては反応時間が短く(facilitation)、低頻度語に対しては反応時間が長い(inhibition)ことを期待する。

facilitation→負の *t* 値、inhibition→正の *t* 値を暗示

自由度 10 に対して *t* 値－2、facilitation を期待

→*t* 分布の左側を使って－2 かそれより低い *t* 値を観察する確率を計算 **> pt(-2, 10)**

＊*two-tailed test*（両側検定）

・頻度の効果について何も知られていなく、頻度が重要であるか（facilitatory か inhibitory か）検定したい場合、*p* 値は 2 倍になる。

> 2 * pt(-2, 10)　*t* 値は正、負の可能性→分布の両側の確率を合計
t 分布の密度曲線は対称的なので、*t* が－2 より小さい率は、
2 より大きい確率と同じ→両側の確率を合計

・*t* 値が 2 の場合

> 2 * (1-pt(2, 10))　*t* が 2 より大きい値の確率を得るために 1 から pt(2, 10)(2 より小さい値の確率)を引いて、両側検定のため 2 倍

・*t* 値の絶対値を使って、正、負の値の検定

> 2 * (1-pt(abs(-2), 10))

> 2 * (1-pt(abs(2), 10))

・Table 4: 片側検定と両側検定のまとめ

＊統計モデルには基本的な特性がたくさんある。

Crawley (2002:17)の指摘

- 全てのモデルは間違っている。
- 他のものよりもいいモデルがある。
- 正しいモデルを確信をもって知ることは決してできない。
- モデルは単純なほどよい。

※ モデルがデータに合うかどうかをチェックすることが大切である→*model criticism*
検定は非常に小さい *p* 値を算出するかもしれないが、検定が基づいている仮説が間違ったものなら、*p* 値は全く役に立たない。