



仮説検定の理解と再現性【第111回生物統計学】

1 概要

仮説検定は、分析者が立てた仮説が正しいのかを統計学的に検証するための手法であり、臨床試験から結論を導くために必要不可欠な手法です。しかし、仮説検定の計算は、統計ソフトなどで簡単に求められることから、理屈をきちんと理解せず使用している人が多く、これにより、再現性の危機といった重大な問題を引き起こすことが考えられます。

本稿では、仮説検定の理解が不十分なことによって生じる再現性の危機についてご紹介します。

2 再現性の危機

再現性とは、同条件で試験を行った際、同一の結果が得られることを指しており、この再現性があることによって、試験の妥当性が判断できます。そのため、再現性がない場合、試験結果から導き出した結論が誤っている可能性があり、これを再現性の危機といいます。

再現性が低くなる要因として、仮説検定の使い方が挙げられます。まず、仮説検定の結果として P 値が設定した有意水準を下回れば、統計的に有意として帰無仮説を棄却し、対立仮説を採択できます。そのため、P 値のみに着目して解析結果に目を通してしまいますが、事後のデータを取捨選択することで P 値を操作することができてしまいます。仮説検定では、第一種の過誤を制御しているという点が重要ですが、p-hacking をすることにより、第一種の過誤を起こす確率が大きくなってしまうため、再現性が低い研究結果が報告されてしまいます。

3 P-hacking の例

前項でも述べた通り、p-hacking とは、P 値を操作して、結果を都合のいいように解釈することを指しています。そのため、どのような状況が p-hacking に該当するのか、例をいくつか紹介します。

1. 試験終了後、P 値が設定した有意水準を下回らなかったため、有意差が出る人数になるまでサンプルサイズを追加した。
2. 複数の実験を繰り返し、有意差があった組み合わせのみを仮説検定の結果として採用した。
3. 試験終了後、得られたデータを有意差が出るように書き換えた。
4. あらかじめ設定していた解析手法以外で有意差が出たため、そちらの解析手法で論文を投稿した。



4 まとめ

試験において、再現性の高さは、試験結果が科学的に正しいことを証明する重要な要素です。このため、仮説検定や意図せず再現性を低下させる p-hackingなどをよく理解し、解析を行う必要があります。

5 参考文献

- 阿部 真人. 統計学入門. ソシム株式会社. 2022: 232-260

ヒト臨床試験（ヒト試験）で得られる結果は、様々な誤差を含んでいます。この誤差を小さくすることで介入効果を増大させることができます。オルトメディコは、多分野の専門家を有するため、様々なアプローチにより誤差を最小化する試験運営が可能です。引き続き、皆様にご満足いただけるような高品質なヒト試験を提供させていただきますので、今後ともどうぞ宜しくお願い申し上げます。