

視覚評価の基礎

ー統計学的検定の概念と検定手法の選択ー

八戸市立市民病院 医療技術局放射線科 ○佐藤 匠(Satoh Takumi)

【はじめに】

統計学的検定とは、母集団に関する仮説が統計学的に成り立つか否か、標本のデータを用いて判断することとされている。統計学的検定を実施することにより、データ分析の結果の正当性を判断することができる。また、少ないデータ量で結果の正当性を確認することができるというのが統計学的検定のメリットとして挙げられる。

臨床研究や論文投稿を行うにあたり、統計学的検定によるデータ分析、評価は必須とされている。しかし、統計学的検定手法の選択に際し、どの検定手法を選択すべきなのか悩ましいといった方が多いのではないと思われる。

本稿では、統計学的検定の概念から検定手法の選択手順、解析結果の解釈について、統計学的検定の実例を用いながら解説する。

【統計学的検定手法の概念】

統計学的検定は、まず対立仮説と帰無仮説を立てるところから始まる。介入群とコントロール群の効果に差があるという仮説を対立仮説、効果に差が出ないという仮説を帰無仮説とした場合、帰無仮説を否定できるかどうかを証明することで、対立仮説を立証するというのが統計学的検定の考え方となる。

次に、検定を行うにあたり、標本データの抽出が必要となる。母集団の中から標本データをランダムに抽出し、介入群とコントロール群にランダムに割付をする。この標本間に介入による差がないと仮定するのが帰無仮説となるが、帰無仮説が正しいと仮定した場合、標本間の誤差は、抽出、割付のランダム性による偶然の誤差となる。このランダム性のみによって、差が生じる確率がP値と定義される。このP値の定義より、統計学的検定の標本抽出は、ランダムに抽出、割付されていることが前提となる。

標本データの抽出後、統計解析ソフトにて検定統計量の算出、P値の計算を行う。設定した有意水準と計算されたP値を比較し、P値が低ければ帰無仮説は棄却され、統計学的に有意な差が認められるということになる。なお、有意水準は慣例的

に5%と設定されることが多い。

【統計学的検定手法の選択】

統計学的検定手法の選択にあたり、重要なポイントを下記に記載する。

- ①変数の種類（連続変数、順序変数、名義変数等）。
- ②正規分布に従うか否か。
- ③パラメトリックデータかノンパラメトリックデータか。
- ④群の数（2群か3群以上か）。
- ⑤対応の有無。
- ⑥等分散性の有無。

統計学的検定手法の選択にあたり、最も重要になってくるのが解析の対象となるデータがパラメトリックデータなのかノンパラメトリックデータなのかを正しく判定することである。パラメトリックデータかノンパラメトリックデータかの判定手法についてFig.1に示す。連続変数以外の変数についてはノンパラメトリックデータとなる。連続変数である場合は、そのデータが正規分布に従うかどうかを評価する必要があり、正規性の検定を行う。正規性の検定の結果、帰無仮説が棄却された場合は正規分布に従わないデータとなるため、ノンパラメトリックデータとして取り扱い、棄却されなかった場合はパラメトリックデータとして取り扱うことができる。

パラメトリックデータかノンパラメトリックデータかが判定できたら、比較する群の数、対応の有無、等分散性の有無によって、検定手法を選択してい

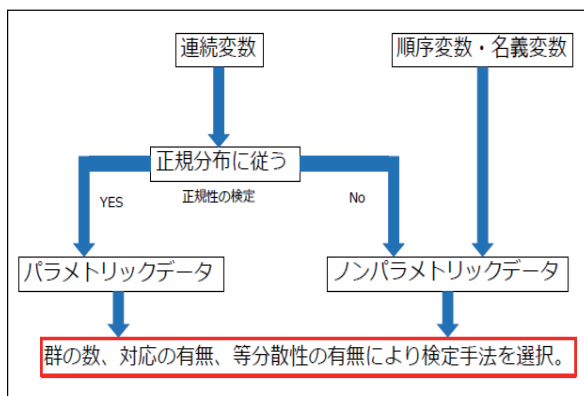


Fig.1 パラメトリックデータとノンパラメトリックデータの判定

く。2群間比較の場合の検定手法の選択方法について、Fig.2に示す。

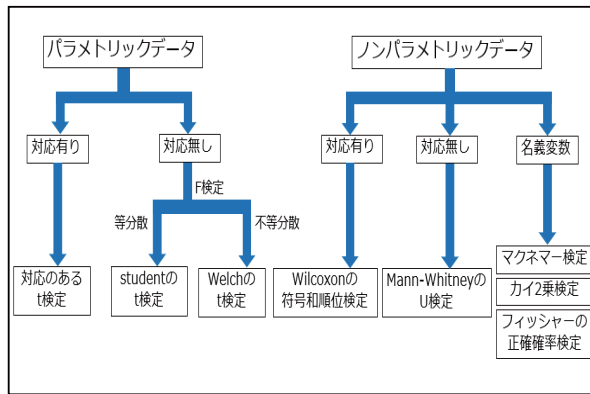


Fig.2 2群間の比較

パラメトリックデータでは、対応の有無で検定手法が異なり、対応無しのデータでは等分散性についても検定を行う必要がある。ノンパラメトリックデータでは、名義変数の場合とそれ以外とで検定手法が異なる。対応の有無で検定手法が異なるのはパラメトリックデータと同様である。

3群以上の多群間比較の場合の検定手法の選択方法について、Fig.3、Fig.4に示す。差の検定の選択方法については、2群間の比較の場合と同様の流れとなる。Fig.3の多群間比較では、対象となる群全体における有意差検定となり、群間毎の有意差の評価をすることはできない。群間毎の有

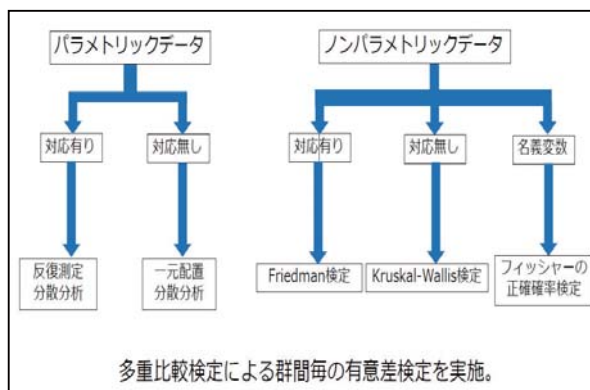


Fig.3 3群間以上の検定（差の検定）

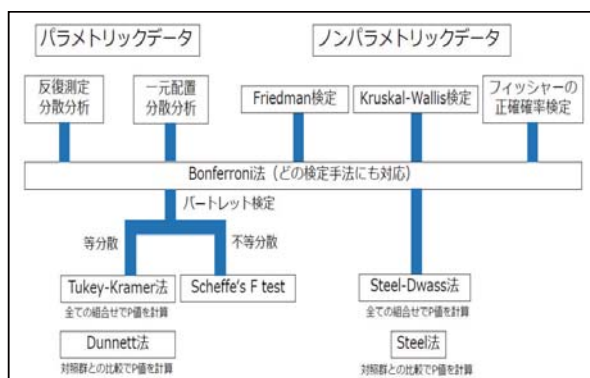


Fig.4 多重比較検定

意差検定を行う場合は、多重比較検定による検定を行う必要がある。多重比較検定における検定手法の選択方法は、差の検定の際に用いた検定方法に対応させる形で検定手法を選択する (Fig.4)。Bonferroni法はどの検定手法に対しても対応しているが、比較する群の数が多くなるほどP値の補正が大きくなり、検出力が低下する性質があるため注意が必要である。なお、多群間の比較において、2群間検定を群間毎に繰り返してはならない。これにより標本全体の α エラーが増大し、有意になる確率が大きくなってしまいうためである。多重比較検定は、原理的には2群間検定を繰り返すことにより生じる α エラーの増大を補正するための手法となる。

【統計学的検定の実例】

統計学的検定の実例として、撮影線量毎の模擬陰影の検出能評価への影響について、統計学的検定を用いて評価を行った。模擬陰影を設置した胸部ファントムを4種類の撮影線量 (0.5 mAs、1.0 mAs、1.6 mAs、3.2 mAs) にて撮影し、取得した胸部画像について視覚評価を実施し、ROC解析にてAUCを算出した。視覚評価の試料数は、それぞれの撮影条件にて陽性20例、陰性20例とし、11名の観察者にて評価を行った。統計学的検定を用いることにより、撮影条件の最適化を検討することを目的とした。統計解析ソフトには、statcel ver5を使用した。

上記の結果得られた撮影線量毎、視覚評価者毎のAUC値をFig.5に示す。本例における統計学的検定の手法について、Fig.1～Fig.4のフローチャートを用いて選択する。AUC値は連続変数であるため、正規性の検定によりパラメトリックデータかノンパラメトリックデータかを判定する。正規性の検定結果をFig.6に示す。正規性の検定結果より視覚評価者間のデータにおいて正規性が認められなかった群が2群認められた。よって、本データにおける正規性は棄却され、ノンパラメトリックデータとなる。本例では、比較する撮影線量が4種類であることから、3群間以上の比較検定の手法を用いることとなる。対応の有無については、同じ胸部ファントムに対し撮影線量を変更した試料を比較しているため、対応有りの検定を選択する。よって、Fig.3のフローチャートより、Friedman検定を選択することとなる。群間毎の有意差を評価するため、多重比較検定を行う必要がある。Fig.4より、Friedman検定に対応する多重比較検定はBonferroni法となる。

統計学的検定の結果をFig.7に示す。Friedman検定においてP値は0.009となり、いずれかの撮影線量間で有意な差が認められた。多重比較検定の結果、0.5 mAs-3.2 mAsの線量間において有意な差が認められた。撮影線量毎のAUC値について、Fig.8に示す。中央値の変化としては、1.0 mAsから1.6 mAsにかけて大きな変化が認められたが、この群間において統計学的に有意な差は認められなかった。

1.6 mAs-3.2 mAsの線量間のAUC値に大きな変化は認められず、統計学的な有意差も認められなかったことから1.6 mAs以上の撮影線量は不要と考えられる。0.5 mAs-1.0 mAsの線量間のAUC値にはほとんど差が認められなかった。一方、0.5 mAs-3.2 mAsの線量間で統計学的な有意差を認めていることから、1.0 mAs以下の撮影線量では過小線量となることが考えられ、本検討における線量の適正化は、1.0 mAs-1.6 mAsの範囲で考えるのが妥当であるという結果となった。

本検討においては、視覚評価者間のAUC値への影響が考慮されていなかったが、Friedman検定において視覚評価者間の結果に有意差が認められるという結果となっていたことから、視覚評価者の承認を得た上で、その影響について検討を加える必要がある。また、本検討では、統計学的検定により撮影線量の適正化を図る上でのカットオフラインの検討を行ったが、統計学的検定により算出されるP値は、サンプルサイズやデータのばらつき等複合的な要因で影響を受けるため、統計学的検定の結果のみで撮影線量の適正化を考察することはできない。統計学的検定の結果で線量の適正化の目安を考えることはできるが、その他に物理評価等多方面からの評価が必要であり、その上で臨床的に有意な撮影線量を決定していくことが肝要である。

視覚評価者	0.5mAs	1.0mAs	1.6mAs	3.2mAs
a	0.734	0.782	0.830	0.791
b	0.788	0.720	0.798	0.855
c	0.795	0.554	0.746	0.758
d	0.670	0.530	0.781	0.722
e	0.698	0.780	0.908	0.968
f	0.804	0.799	0.764	0.899
g	0.631	0.675	0.737	0.705
h	0.614	0.618	0.769	0.800
i	0.546	0.615	0.662	0.736
j	0.566	0.620	0.646	0.611
k	0.750	0.816	0.710	0.769

Fig.5 撮影線量毎のAUC値

列間 (撮影線量)		行間 (視覚評価者)	
	P値		P値
0.5mAs	0.265	a	0.432
1.0mAs	0.623	b	0.432
1.6mAs	0.337	c	0.041
3.2mAs	0.337	d	0.266
		e	0.432
		f	0.041
		g	0.432
		h	0.266
		i	0.432
		j	0.149
		k	0.432

Fig.6 正規性の検定結果

Friedman検定 (撮影線量間)		Bonferroni法 (撮影線量間)	
自由度	3		P値
グループの数	4	0.5mAs,1.0mAs	4.738
同順位の数	0	0.5mAs,1.6mAs	0.197
χ^2 値	11.618	0.5mAs,3.2mAs	0.035
P値(上側確率)	0.009	1.0mAs,1.6mAs	0.197
同順位補正 χ^2 値	11.618	1.0mAs,3.2mAs	0.087
同順位補正P値(上側確率)	0.009	1.6mAs,3.2mAs	1.380
$\chi^2(0.95)$	7.815		

Fig.7 統計学的検定の結果

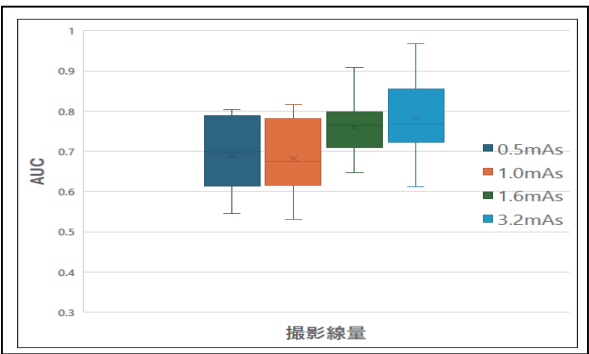


Fig.8 撮影線量毎のAUC値 (箱ひげ図)

【まとめ】

統計学的検定の概要と検定手法の選択手順、検定の実例について記載した。統計学的検定では、対立仮説と帰無仮説を明確にすることと、標本抽出がランダム抽出、ランダム割付であることが非常に重要である。検定手法の選択については、データの性質を正しく理解した上で、フローチャートに則って検定手法を選択していくことで、適切な検定手法が選択できると考えられる。また、統計学的に有意であることと臨床的に有意であることは分けて考えるべきである。臨床的有用性の評価は多角的視点から行われるべきものであり、統計学的な評価はあくまでその中の一つとして認識する必要がある。

【参考文献・図書】

- 1) 柳井 久江: 4steps Excel統計 (第5版) オーエス
エム出版
- 2) 大林 準: 統計検定手法の選び方ー基本編ー

天理医学紀要 2022;25 (1) :60-65

- 3) 神田 善伸: EZRでやさしく学ぶ統計学: EBMの
実践から臨床研究まで 改訂3版 東京: 中外医
学社