LIFE2012　2012 年 11 月 2 日－4 日 愛知 (名古屋大学)

人間から取得するデータの統計的評価

Notes on statistical methods for data

対馬栄輝（弘前大学大学院保健学研究科）

Eiki TSUSHIMA, Hirosaki University Graduate School of Health Sciences

Abstract: This paper noted some problems in the statistical hypothesis testing. The immediate implication of significant result for clinical practice is not obvious. The p-values reported do not translate into estimates of the magnitude of the positive outcome of the intervention. However, confidence intervals (CIs) around point estimates (e.g., an estimate of the population mean) indicate the range of values within which the true point estimate probably lies. Effect size (ES) can also be used to express the size of the effect (e.g., difference of the mean, correlation　coefficient). Cls and ES will provides information in an easily understandable form.

Key Words: Statistical hypothesis testing, Confidence intervals, Effect size

1. 人間から取得するデータの特徴

人間から取得するデータに限ったことではないが，データは，

データ＝真の値+誤差

によって構成されると考える．この誤差の内訳は，被検者の誤差，検者の誤差，機器の誤差など，混在している状態となる．

誤差は，統計的な扱いの観点から偶然誤差と系統誤差に分けられる．

1-1 偶然誤差

偶然誤差とは，データを測定する際に発生する必然的な誤差で，無意識のうちに付随する．真の値に対して無意識にプラス方向，マイナス方向の誤差がランダムに付随するために，その誤差成分は平均 0 の正規分布に従うと考えられる．

偶然誤差は幸いにも，正規分布に従うことと，大小様々ランダムに発生すると考えられるので，統計的な手法を使った対策が可能である．

1-2 系統誤差

系統誤差は偶然誤差とは異なり，プラス方向またはマイナス方向へ偏って付随する誤差（特にバイアスともいう）のことである．系統誤差は，常に一定量発生するときもあるし，条件や環境によって段階的に増えたり減ったりすることもある．系統誤差は，定量的に存在するために，統計的な手法を使っても対策は不可能である．

例えば，立ち座りに快適な椅子の高さを決める実験をするとする. 調べたい要因は, 膝の角度 3 条件（80deg., 90deg., 100deg.）と，座面の奥行き 3 条件（20cm,30cm,40cm）である．膝の角度を 80－90－100deg.の順に，また座面の奥行きを 30－40－50cm の順に立ち座りを繰り返して測定する．もし下肢の疲労が現れるとすれば，ほとんどの被検者は，100deg.かつ 40cm の条件が，最も疲労の影響を受けて「立ち上がりにくい」と答えるに違いない．逆に 30cm かつ 80deg.の条件が最も立ち上がりやすいと答えるだろう．

これは疲労という系統誤差が入り込む例であるが，例えばラテン方格を利用して実験順序を計画的にすると，誤差を各条件へ均等に配分することが可能となる（Table 1）．A という人にはまず 80deg.で 30－40－50cm の順に，次に 90deg.で 50－30－40cm, 最後に 100deg.で 40－50－30cm と立ち座りさせる．今度は B という人に，90deg.の 50－30－40cm から始め，次に 100deg.の…，C という人には 100deg.から始めて，次に 80deg.の…，という手順で，角度・座面の奥行きの測定順番が均等となるようにバランスをとれば，疲労の影響が各条件へ均等に入るはずである．

Table 1　A example of the use of Latin square design

	30cm	40cm	50cm
80 deg.	1st	2nd	3rd
90 deg.	2nd	3rd	1st
100 deg.	3rd	1st	2nd

このように測定の順序を考慮して，系統誤差を各条件へ均等配分するような工夫を局所管理という．

厳密に統制された機器を対象として実験を行う場合は大きな問題とならないだろうが, 人間を対象とした研究では, このような局所管理を考慮する必要がある．

2. データの特性値（代表値と散布度）

正規分布に従うデータに対する代表値としては，平均と標準偏差 sd を用いる. データが正規分布に従うときは平均と中央値は一致するために，どちらを用いても良いのであるが，数理的に良い性質をもった平均を優先する．標準偏差は，平均から求められるために，常に一緒に用いる．

データが正規分布に従わないときは平均と中央値の一致しない場合が多い（Fig.1）．従って，より分布のゆがみに影響を受けにくい中央値と四分位範囲を用いる．

データが正規分布するか否かは，シャピロ・ウイルク検

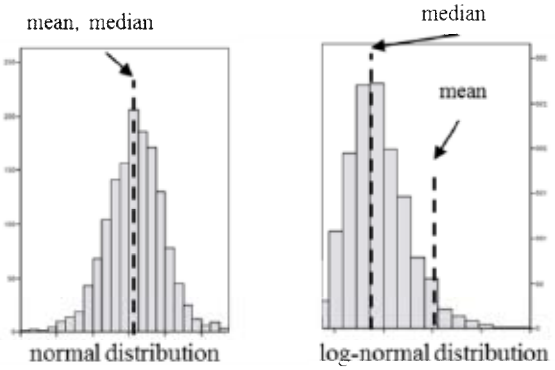


Fig. 1　parameters of data distribution

定による判断が妥当である．シャピロ・ウイルク検定は，オープンソースとして提供されている R[(1)]でも計算可能である．

3. 統計的検定結果の解釈における注意点

統計的検定の手法を選ぶためには，成書[(2)]を参考とする方が簡単で良い．現在では多くの良書があるので，さほど困らないであろう．

検定を行って，算出された有意確率 p が，p<0.05 もしくは p<0.01 のとき，有意に差があるとか，有意な相関があると判定する．しかし，この有意という解釈に誤解がないようにしなければならない．いかなる検定も標本の大きさ n が大きくなると，推定精度が高まるゆえに有意となる性質がある．これを踏まえた上で，以下にその注意点を挙げる．

3-1 検定結果が有意なときは，その程度を評価する

例えば差の検定の結果で考えると，p<0.05 という結果よりは p<0.01 の方が確率は低い．しかし，「p<0.05 より p<0.01 の方が，差が大きい」という意味ではない．

つまり，平均や中央値の差が大きかろうが小さかろうが p の大きさとは無関係である．統計的に有意であるということは，その差が 0 という事象が否定されただけであって，差が 0 以外の 0.000…1 のように小さい差から無限大までの大きな差の何れかという意味である．つまり，p が小さいほど差が 0 である可能性が小さいというだけであって，差の程度は不明なのである．これは，他のあらゆる検定に共通の原理である．

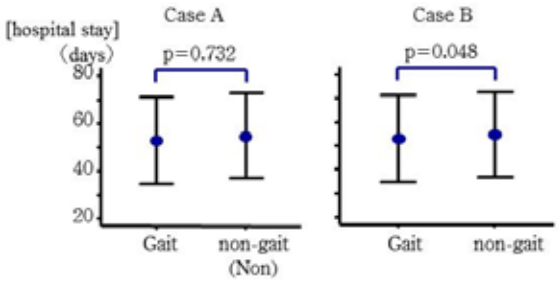
従って，p<0.05 より p<0.01 の方が差は大きいとは限らず，単純に「p<0.05 よりは p<0.01 の方が差のある可能性は確実なだけ」という解釈が正しい．ゆえに，差の程度を表す他の情報が必要となる．

3-2 差の程度を表す信頼区間と効果量

差の程度は，95%信頼区間 confidence interval（95%CI）や効果量 effect size（ES）を参照する．95%CI は，ほとんどの統計ソフトで出力される．ES は，Web にて簡単に計算できる Excel ファイル[(3)]が配布されている．

例えば，Fig.2 には，ある疾患で入院した患者を対象として，退院時に歩行が可能であった群と，歩行が不可能であった群の平均入院日数の差を検定した 2 つの例 A，B を挙げている．例 A も例 B も 2 群の平均は同じで，もちろん平均差の値も同じである．しかし，例 A は有意な差は認められず，例 B は有意な差がある．例 B は n が大きいのである．これらの ES を求めてみると，両者とも 0.1 で同じとなる．95%CI は，例 A が－8.77～12.39 日，例 B が 0.02～3.6 日であった．

95%CI とは，歩行可能群と歩行不可能群の母平均の差（真の平均差）が「95%の可能性でその間にある」と推測する範囲を表す．例 A の場合は母平均差が-8.77～12.39 日なので，今後対象者数を多くすると，「95%の可能性で歩行不可能群よりも歩行可能群が 12.39 日長くなる」から「歩行不可能群よりも歩行可能群が 8.77 日短くなる」範囲をとる，と推測できる．例 B の場合は，母集団平均の差が 0.02～3.6 日，すなわち 95%の可能性で「歩行不可能群よりも歩行可能群が 0.02～3.6 日長くなる」範囲をとる，と推測される．例 A では，歩行可能群と歩行不可能群の母集団平均差が，95%で逆転する可能性もあり，0 日にもなり得るから，差があるといいきれない．例 B の場合は 95%の可能性で，最小 0.02 日から最大で 3.6 日の差があると推測され，辛うじて平均差＝0 ではないので有意となっている．しかしながら入院期間の平均差が最大でも 3.6 日の差というのは，患者または病院の事情によって変動する範囲だと判断されるなら，これは有意な差があっても，実質的に有効な差とは考えられない．このように 95%CI を利用して差の程度を推測することができる．



		N	Mean (days)	sd	P	ES	95%CI(days) lower	95%CI(days) upper
A	Gait	19	52.9	18.8	0.73	0.1	-8.87	12.39
	Non	34	54.7	18.2				
B	Gait	608	52.9	18.4	0.04	0.1	0.02	3.60
	Non	1,088	54.7	17.9				

Fig. 2　Two examples of difference averages

ES は n の大きさ，バラツキ等を調整した上で，データの差の程度を表すものである．95%CI と異なるのは，推定の意味が無いことである．相関で例えると，ES は相関係数そのものである．これと照らし合わせて考えると，ES の意味が理解できるであろう．

Fig.2 の例 B は有意差があるものの，結局，差の程度は 0.1 で例 A と同等と判断できる．

3-3 検定結果が有意でなかった場合の対処（検出力分析）

検定を行って，有意ではなかったとき，n が少ないという問題があり得る．

そこで，検定に必要な最低限の n 数を上回っているか検討する必要がある．この計算には，①検出力，②有意水準 p，③ES，④n のうち，3 つが決まれば残り 1 つが決まるという性質を利用する．①～③を決めると検定に必要な最低の n が求まる．これらを求める手順を検出力分析 power analysis という．

検出力分析は統計ソフトを用いる方が良い．幸いにも計算可能なフリーソフトがいくつかあるので，それを利用するのが手っ取り早い．たとえばフリーソフトの G*power[(4)]は，簡単な手法に限って日本語の解説書も Web 上で配布[(5)]している．

参考文献

(1) http://www.hs.hirosaki-u.ac.jp/~pteiki/research/stat/S/
(2) 対馬栄輝：SPSSで学ぶ医療系データ解析，東京図書，2007.
(3) http://www.mizumot.com/stats/effectsize.xls
(4) http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/download-and-register
(5) http://www.hs.hirosaki-u.ac.jp/~pteiki/research/stat/gpower.ppt