解説【仮説検定とベイズ統計】

● 「統計的に有意」は仮説検定で用いられている用語である．放射線リスクの研究では，β=0(リスク上昇なし)を仮説にとりβ>0(リスク上昇あり)を対立仮説にとって仮説検定をおこなうのが一般的である．(線形モデルによる回帰分析の過剰相対リスク係数がβである考えて話を続ける.） 検定の目的は仮説β=0 がデータから否定されるかどうかを調べることである．そのためにβ=0 の母集団から抽出した標本の分布を考える．この標本のひとつがわれわれの手にしているデータである．標本の分布はβがマイナス無限大からプラス無限大にまで広がっている．したがってデータのβがどんなに大きくても母集団の仮説 β=0 を否定することはできない. そこで,ある程度おおきな値 c を決めておいて, β>c であれば仮説を棄却することにする．棄却できるのではなく棄却することにするのである．仮説を「棄却する」のは取り決めについての表現であって，仮説を「否定する」ということはいえない．仮説が棄却されたとき「対立仮説 β>0 が統計的に有意である」という仮説検定独特の表現を使う．c は標本の分布上でβ>c の確率がαになるように決める．αを有意水準という．標本の分布上でデータのβの値より大きい区間の確率を p 値という. したがってデータのβと母集団のβのへだたりが大きくなるほど p 値は小さくなる. 放射線リスクの研究では 0.05 という数値が有意水準としてよく使われいる． p<0.05 のとき仮説を棄却するのである．仮説が棄却されたとしても仮説がなお正しい確率は 5%残っていることは承知しておく必要がある．(以上は片側検定の説明であり，β≠0 を対立仮説として検定をする場合もある．この検定を両側検定という．)

信頼区間も標本の分布をもとに考える. 信頼区間の目的は母集団のβが確実に含まれる範囲を知ることである．この場合もデータから母集団のβの確実な範囲を決めることはできない. そこである確率で母集団のβが存在する範囲を決めることで満足することにする. これが信頼区間である. たとえば 95%信頼区間は 100 の標本をえたとき 95 の標本ではそれが母集団のβが存在する範囲を正しく与えるが, 残る 5 標本では正しい区間を与えないことを容認するということである．信頼区間と仮説検定は密接な関係にあって, 信頼区間をもとに仮説検定の結果を知ることができる．仮説 β=0 が信頼区間の下限より下に位置すれば仮説は棄却され，対立仮説 β>0 が統計的に有意になる．95%の信頼区間は上

限より上が 2.5%，下限より下が 2.5%と考えてよいだろうから対応するのは有意水準 2.5%の片側検定である.

● 仮説検定は標本の分布をもとにした分析法であるが，これと違って直接，βの値を推定する統計学がある．ベイズ統計学である．これは最初からβの確率分布を考え一貫してβの確率分布を扱う. したがってβ>0(リスク上昇)の確率を直接求めることができる. 原爆被爆者の固形がん死亡率の分析で 100mSv 以下でのリスク上昇が統計的に有意でなかったという論文があるが，これはリスク上昇がなかったことを意味しない．リスク上昇が統計的有意でなかったこととリスク上昇がないこととは全く別ものなのである. しかし，ベイズ統計学では，そのデータを使ってリスク上昇（β>0）の確率が 95.3%(概算)であることが導かれる．リスク上昇なしの確率はわずかに 4.7%である．100mSv 以下でもリスク上昇がはっきり見えていることになる. リスク上昇が統計的に有意になるのはベイズ統計でいえば β>0 の確率が 97.5%以上のときである．仮説検定がリスク評価にあたっていかに鈍感であるかがわかるであろう.

● 米国統計学会が声明で警告したように, 仮説検定は研究の世界に大きな弊害をもたらしている．ここで声明を詳しく解説するわけにいかないので，放射線リスクの研究で特に現れている弊害についてだけ指摘しておく．第一に，仮説検定はデータがほんとうはもっている情報のおおくをむざむざ捨てている．天気予報の例でいえば,「あすの降水確率は 20%」という予測が可能なのに，仮説検定だと「あす雨が降る」という仮説が棄却されるかどうかしか問題にしない．結果は○か×しかない．第二に，仮説検定は仮説をどう選ぶかで結論の印象が大きく変わるという欠陥をもっている．100 ミリシーベル以下でがんリスクの増加が棄却されていないことを先に述べた．この結果から「がんの増加を示す証拠は見つかっていない」と人々に宣伝する人々がいる．しかし，同じ仮説検定を使ってがんリスクが被曝量に比例して増加するという仮説で検定すれば, この仮説も棄却されない．そうすると，この結果から,「がんリスクが被曝量に比例して増加することを否定する証拠はない」ということになる．このように同じデータを使っても大きく印象の違う表現が可能になってしまう.

仮説検定はリスクの検知に鈍感であるとともに, 仮説のとり方で自説に有利な誘導ができてしまうという欠陥をもっている. ベイズ統計であればどちらの仮説が妥当であるかを数値的に示すことができるのである.

- より詳しく正確に仮説検定とベイズ統計学を知りたい人はそれぞれの教科書をみつけて勉強してほしい.

（文責　永井宏幸）