

○きちんと統計がほしい。

今回トラスクリプトーム（細胞内の転写産物 mRNA の総量。細胞ある変化が起こると増減するし、体中で同じとは限らない）を調べているときに、石田さんというライターさんのブログに、この変化は、喫煙所のクラスター化と関係か、「喫煙所」が「新型コロナ」クラスター発生源にという石田雅彦さんというライター（3/2(月)14:26）というのを見つけた。そこで、これを自分なりの切り口で、なぜ、今の保健所検査のやりかたが納得できないかという別の側面にふれてみよう。

○例えば、100人のヒトがいて、タバコをすっていて、コロナになった＝たばこころが30人、タバコをすっていてもコロナにならなかった＝きんころ5人、タバコを吸わながコロナになっ＝たばなしが10人、タバコをすわないし、コロナにもならない＝きんなしが60人の計100人いるとする。これを表にすると

実数	コロナ+(ころ)	コロナ-(なし)	横計
喫煙(たば)	30	10	40
非喫煙(きん)	10	50	60
縦計	40	60	100

これでは、喫煙と非喫煙は 30 : 10 = 3 : 1 だね。

この表をもとにそうなる割合＝期待度をだしてみる。簡単は場合の数からでる確率で 全員は100  
 コロナ群 40 たば群も 40 たばは全100の内 30+10で40で、なった(たばこころ)が30と言う分けだから、仮定：たばこころななりやすいことが関係なしとする場合、この数字の期待度数を求めると、

たばこころ  $40 \times 40 / 100 = 16$  きんころ  $60 \times 40 / 100 = 24$  たばなし  $40 \times 60 / 100 = 24$   
 きんなし  $60 \times 60 / 100 = 36$ 、表にすると、

期待値	コロナ+(ころ)	コロナ-(なし)	横計
喫煙(たば)	16	24	40
非喫煙(きん)	24	36	60
縦計	40	60	100

これが正しいかどうかカイ二乗検定という方法をつかってみると、

たばこころ  $(16-30)^2 / 16 = 12.25$

きんころ  $(24-10)^2 / 24 = 8.17$

たばなし  $(24-10)^2 / 24 = 8.17$

きんなし  $(36-50)^2 / 36 = 5.44$

合計が 30.028

という数字になる。同じく表にすると、自由度1で

カイ二乗値	コロナ+(ころ)	コロナ-(なし)	横計
喫煙(たば)	12.25	8.167	
非喫煙(きん)	8.167	5.444	
縦計			34.028
		null hypothesis has denied.	
		->OK	

で、カイ二乗値（上記期待値の総計で、期待値と実測の分布の正規分布とのズレ）は34.03となるので、この値では、統計上p値（タバコとコロナが無関係する仮説イコール帰無仮説が成立したときの確率）が、一般的には有意水準（ここでは<0.05つまり5%以下）値が5.44なので十分大きいから「仮定は否定された」から、逆にタバコとコロナは密接な関係あるが、となるというわけだ。

この表のなかの値は分母が大きければ大きいほど信頼性がたかまるのはだれがみても分かる。しかもコロナである検査を症状にわけてしたりしなかったりすると、コロナの検査の確度自体が定まらない今、全数検査しないと統計もとれないから、こういう統計も作れないし、本当にこまる。真実が分からなくなる。日本のやり方は、本当にマズい。

この表は、いろいろできる。まずみんな知りたがることの一つに検査の信頼度（感受性と特異度の検定）がある。つまり、ころけんさ+（真陽性）、ころけんさ-（偽陰性）  
 ころなし検査+（疑陽性）、ころなしけんさ-（真陰性）

真陽性＝テスト感受性、真陰性＝特異度の2分割表が検査の正確さと直接関係するから、下に書いたような迅速テストができたとして、実際に使ったときどうするのだろうか。なんとかしてほしい。