PAS Kara News（152）

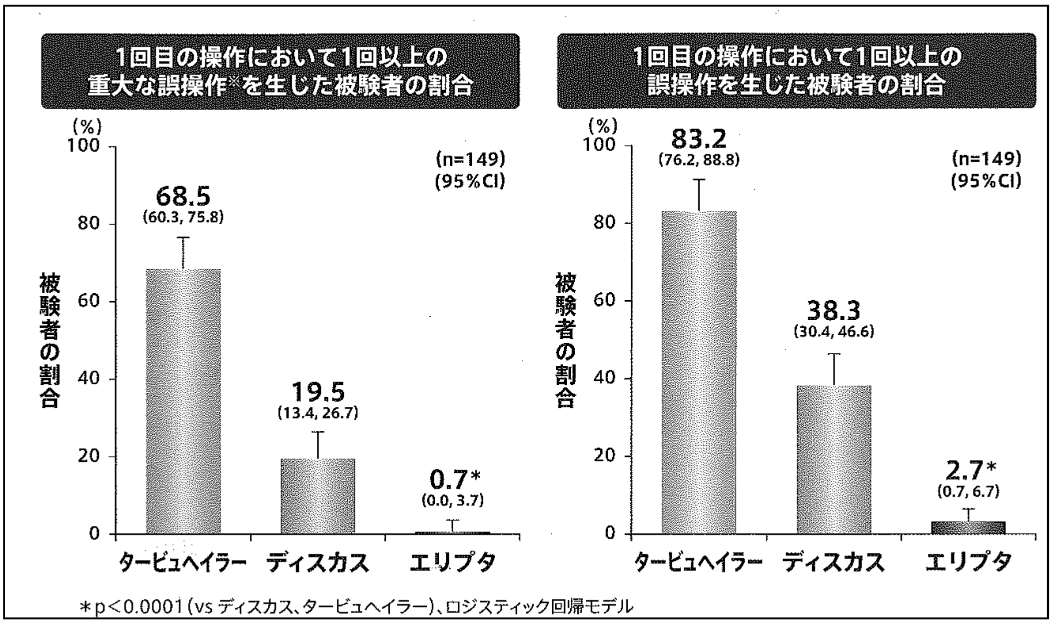
平成 27 年 5 月 29 日
企画編集：足立博一
www.adachipas.com

ロジステッィク回帰分析

「薬の評価に役立つような分かりやすい統計学の話ができないですか？」とある薬剤師から言われて、その道の専門でもないので難しいなと思いつつも、その気になって勉強している最中に、ある薬局さんのメーカー学習会に参加した折に、下記のスライドが示されました。

エリプタという吸入器は１回操作で吸入ができるので誤操作をする割合が減るため、吸入薬のコンプライアンスの向上に役立つという根拠データを示したものでした。

１回目の操作で誤操作をする人が他の吸入器と比べて有意な差をもって少ないことが示されています。左のグラフでみると、１００人の人が初めて吸入器を使う時に間違えてしまう人がディスカスで約２０人いるのに対して、エリプタだと１人いるかいないかのレベルだという訳です。



統計学的な根拠としては、ディスカスとエリプタの棒グラフに示されている９５％信頼区間に重複部分がないので有意に差があると言えるわけです。

そこで、ある薬剤師が質問してきました。９５％信頼区間というのは、標本の平均値の±αの区間に母平均(真の平均値)の入る確率が９５％になるものなのに、タービュヘイラーの９５％信頼区間が６０．３から７５．８となっていて足して２で割ると標本平均は６８．１になるはず。それなのに６８．５になって僅かだけれども違っている。同様にディスカスは２０．１なのに１９．５だし、エリプタも１．９のところが０．７になっている。どうして区間の真ん中の値が平均の値になっていないのですか？「そんな事が私に分かるものか？」とは言えないのが私の悪い性分で、「ちょっと調べてみるわ」と時間をもらい、何とか誤魔化してでも回答できないかというのが今回のお話です。

今回の統計処理で肝になるのは、おそらくグラフの欄外に記載されている「＊p<0.0001（vs ディスカス、タービュヘイラー)、**ロジスティック回帰モデル**」と思われます。

①ｐとは

ｐは確率を意味する probability のｐで、ある出来事が起きる割合を示します。

ここではエリプタとディスカス、エリプタとタービュヘイラーの誤操作の違い、つまりエリプタが優位でない割合は同じ試験を１万回やってもなお１回も起きないようなレベルだという事を示しており、エリプタが圧倒的に優れていますよという意味になります。

②ロジスティック回帰モデル

回帰モデルというのは、「ある要素」と「ある出来事」に１：１の関係がある時、たとえば**「喫煙」**と**「動脈硬化」**の発生率に１：１の関係がある時に

動脈硬化発生率＝β＋α・喫煙率　（グラフにした時、β；切片、α；傾き）　１式

という回帰方程式が求められたとします。これを**単回帰分析**と呼びます。要素が喫煙という１個だけなので**単回帰**と呼びます。喫煙具合によって動脈硬化発生の**予測が可能**になる式になります。

しかし、世の中、起こる現象は、それほど単純ではありません。たとえば動脈硬化発生率に喫煙ばかりでなく、**「年齢」**や**「体脂肪率」**が関与していたとしたらどうなるか（以下、仮想例です）。

動脈硬化発生率＝β＋α1・喫煙率＋α2・年齢＋α3・体脂肪率　２式

という回帰方程式が求められます。重なる要素がでてきたので、これを**重回帰分析**と呼びます。

しかし、世の中は、さらに複雑で重回帰分析のような回帰方程式で収まらない現象もでてきます。それを解決する手法として**ロジスティック回帰分析**があり、例えば下記のような式になります。

$$\text{動脈硬化発生率} = \frac{1}{1 + e^{-(\beta+\alpha 1\cdot\text{喫煙率}+\alpha 2\cdot\text{年齢}+\alpha 3\cdot\text{体脂肪率}+\alpha 4\cdot\text{性別})}} \qquad \text{３式}$$

２式までは線形回帰と呼ぶものでしたが、３式は非線形回帰となっています。とはいうものの複雑怪奇な式でさっぱり意味が分かりません。**こういう式になるのだ**という理解でもよいかと・・・。

さて、このロジスティィク回帰分析の発生率の式は**オッズ比(odds 比)と密接な関係**があります。**オッズ**とは「起きる確率(P) 」を「起こらない確率(1-P)」で割ったものです。**オッズ比**とは「あるオッズ」を「他のオッズ」で割った値で、ある現象とその要因の関係の強さを現わす値になります(今回この辺りも、そういうものだとさらりと読み飛ばしてください)。

あくまで例え話になりますが、３式の例で、**喫煙歴と体脂肪率が同じ**で、**５０歳の男性**と**６０歳の男性**で**動脈硬化発生率は何倍違うか**という問題がでたとします。

今回の例では、６０歳のオッズと５０歳のオッズの比が、実は３式で計算ができるので、たとえばオッズ比が１．７倍となったとすると、年齢が１０年上がると動脈硬化発生のオッズが１．７倍に上がると表現します。

オッズ比でも「１．７倍という**平均値**」に対する**９５％信頼区間の計算式**が設定されています。ここでは式は示しませんが、ネイピア数と呼ばれる自然対数の底**「e」の何乗の範囲**で示されます。

たとえば、**途中の計算式**で　1.033±0.248 と±の幅が等しく出ますが、**e に関連付け**すると

$e^{0.785}$（下限）～$e^{1.033}$（平均値）～$e^{1.281}$(上限)　となり、これらを実数になおすと

2.192（下限）～<u>2.810（平均値）</u>～3.600（上限）　となります。

この下限と上限の単純な算術平均は<u>2.896</u>となりますから、<u>2.810</u>とは合いません。

【まとめ】

今回のエリプタのロジスティック回帰モデルはオッズ比では表現されていませんが、**ロジステッィク回帰分析にした時点で、数式には対数計算などが入ってきて、単純な算術平均等が通用しなくなる**と思っておけば、今回の問題はとりあえず解決するのではないでしょうか？？

質問者の標本平均±αという形式の９５％信頼区間は、検討する対象集団の分布が、**正規分布**や **t 分布**をする時になります。　（終わり）

参考資料：わかりやすい薬学系の統計学入門、講談社(2015 年)