

ロジスティック回帰入門 (2)

FooQoo

2020 年 1 月 24 日

1 概要

調査結果を解釈する指標として、「リスク比」と「オッズ比」がある。本稿では、これらの指標についての理解、その違い、有意差の算出方法を紹介する。

2 分割表

リスク比やオッズ比、そして有意差を算出するためには、分割表が必要になる。表 1 は 10 人の患者について調べたものである。このデータに対して、不整脈が喫煙者と非喫煙者を比較した時、両者に差があるか確認する。

データを読み取ると、喫煙者において不整脈の割合は 60% で、非喫煙者における不整脈のある人の割合は 20% であることがわかる。この結果を表にしたものを分割表 (contingency table) と呼び、表 2 として表す。原因と結果の関係を調べるために分割表を作る場合、行に原因、列に結果の項目を置く。

3 リスク比

表 2 の横軸で割って得られた割合はリスクである。このケースでは、不整脈になる危険や恐れが喫煙の有無によって、どの程度あるのかがわかる。次に、リスクを喫煙者と非喫煙者でそれぞれ計算する。喫煙者が不整脈となるリスクは $3/5$ で 60%、非喫煙者が不整脈となるリスクは 20% である。リスクの差に着目すると、リ

表 1 データ例

患者 No.	喫煙有無	不整脈有無
1	0	0

表 2 分割表

喫煙有無	不整脈有無		横計	割合
	ある	ない		
喫煙	3	1	5	60%
非喫煙	1	4	5	20%

スクは喫煙者が非喫煙者を 40% を上回っているということがわかる。喫煙者のリスクを非喫煙者のリスクで割ると、その比は 3 : 1 となる。この値をリスク比 (Risk Ratio) と呼ぶ。この場合では、リスク比 3 より、「喫煙者が不整脈となるリスクは非喫煙者に比べ 3 倍である」と解釈できる。

4 オッズ比

簡単な例によってオッズ比を紹介する。あるゲームを 5 セットを 1 回として計 2 回やったとする。表 3 は 1 回目と 2 回目のゲームの勝敗の成績を示したものである。勝率の比、すなわちリスク比は 3 になる。リスク比から、1 回目の勝率は 2 回目比 3 倍であり、勝率をゲームの強さと考え、1 回目のゲームの強さは 2 回目比 3 倍強いといえる。

次にこの表を利用して、オッズ比について説明する。1 回目の勝数を 2 回目の勝数で割った値をオッズ (Odds) と呼ぶ。そして、1 回目の負数を 2 回目の負数で割った値もオッズと呼ぶ。勝数に着目すると、1 回目と 2 回目で勝数のオッズは 3 です。負数に着目すると、1 回目と 2 回目で負数のオッズは 0.5 となる。そしてオッズ比は、勝数オッズと負数オッズの比となる。この場合のオッズ比は 6 となる。ここで注意するのは、1 回目のゲームの強さが 2 回目より 6 倍とすることは間違いであることである。このような勝率の比較はリスク比でしかできない。オッズ比でわかるのは、結果に対する影響要因であるかどうかである。今回得られたオッズ比は、6 と大きい値であるため、不整脈の影響要因と考えてもよい。

先ほどの例では、オッズ比の有用性を説明しにくいので、新たに喫煙の有無、飲酒の有無、性別を取り上げ、どの要因が不整脈の有無に影響を与えているか調べることにする。表 4 にそれぞれの要因についての分割表を示す。リスク比の順位を見ると、喫煙が 6.67、飲酒が 1.67、性別が 1.07 の順となる。そして、オッズ比の順位は、喫煙が 9.97、飲酒が 3.00、性別が 1.14 で順位はリスク比と同一である。この場合はリスク比とオッズ比のどちらを用いても不整脈の影響を把握することができる。一方で研究目的によっては、影響要因の順位がわかればよい場合もあるので、オッズ比が必要となるシーンは多い。ただし、オッズ比を使う場合は有意差検定が必要となる。まとめると、オッズ比は、複数の影響要因の寄与順位を把握することのみに活用でき、リスクの倍率把握には適用できない、ということになる。

リスク比およびオッズ比が 1 を上回る時、正の相関があると呼び、反対に 1 を下回る場合は、逆相関があると呼ぶ。逆相関の場合、リスクの比の解釈がわかりにくくなるため、要因のポジネガを入れ替えると正の相関になり解釈しやすくなる。

5 オッズ比の活用法

代表的な臨床研究として、「コホート研究」と「ケースコントロール研究」がある。

- コホート研究。例えば、不整脈が無い人をランダムに 400 人抽出し、今までに喫煙したことがあるかど

表 3 成績表

勝負回数	ゲーム		ゲーム数	割合
	勝回	負数		
1	3	2	5	60%
2	1	4	5	20%

うかを調査し、その後の2年間において、喫煙の有無別に不整脈が発生したか追跡調査する。調査開始時点では不整脈は発生しておらず、それから2年後に不整脈の発生を調べる。この研究は、2年後の未来へ向かって調べる研究であり、前向きの研究と呼ぶ。

- ケースコントロール研究。不整脈があると診断された200人とランダムに選んだ健常者200人について、過去の過去の有無を調査する。すでに不整脈があると診断された人と健常者がいて、その時点から過去に遡って喫煙をしていたかどうかを調べる。この研究は、過去へ向かって調べる研究であり、後ろ向きの研究と呼ぶ。

これらの研究の違いを、原因と結果という因果関係から見る。上記の例では、喫煙の有無が原因変数で不整脈が結果変数です。コホート研究は、未来の結果変数(不整脈の有無)を調べる研究であり、ケースコントロールは過去の原因変数(喫煙の有無)を調べる研究である。

後ろ向き研究で集めたデータから分割表を作成し、喫煙者が不整脈になるリスクが56%であることがわ

表4 多数ある影響要因の比較

(a) 喫煙と不整脈の分割表

喫煙の有無	不整脈		横計	リスク
	あり	なし		
喫煙	12	3	15	80%
非喫煙	3	7	10	30%
オッズ	4.00	0.43		

(b) 飲酒と不整脈の分割表

飲酒の有無	不整脈		横計	リスク
	あり	なし		
飲酒	10	5	15	66.8%
非飲酒	4	6	10	40.0%
オッズ	2.50	0.83		

(c) 性別と不整脈の分割表

性別	不整脈		横計	リスク
	あり	なし		
男性	8	7	15	53.3%
女性	5	5	10	50.0%
オッズ	1.60	1.40		

(d) リスク比とオッズ比

要因	リスク比	オッズ比
喫煙	2.67	9.33
飲酒	1.67	3.00
性別	1.07	1.14

かった時、一般的に喫煙する人が不整脈になるリスクが5割を超えているとみなしてよいのか？ この事例では、不整脈があると診断された200人とランダムに選んだ健常人者200人について、過去の喫煙の有無を調査した。したがって、全対象者における不整脈のリスクは200/500の50%で調査対象者のサンプリングに依存する。ただし、リスク比の値を順位で検討するだけであれば、リスク比を使用できる。しかし、この制限下でリスク比を用いるならば、最初からオッズ比を使えば良い。以上の理由から、ケースコントロール研究ではオッズ比を使う。

6 オッズ比とロジスティック回帰

表5から、結果に対していくつかの原因が考えられる場合、それぞれの原因が、結果にどの程度影響を及ぼしているか考える。この表を見ただけでは、結果がわかりにくいので、分割表・オッズ比・リスク比を求める。

この結果を解釈する。結果より、不整脈に影響度のもっとも強いのは喫煙の有無、次にギャンブル嗜好である。飲酒の有無は3番目であった。この結果が事実かどうかは、それぞれの要因の関係を確認することによってわかる。表7より以下のことが考察できる。ギャンブル好きは7人中6人が喫煙者、ギャンブル嫌い13人中11人が非喫煙者である。リスク比も5.57と高く、両者に強い相関があると考えられる。飲酒の有無と喫煙の有無のリスク比は1.40、ギャンブル嗜好と飲酒の有無のリスク比は1.86とそれほど大きな値ではなく、

表5 ロジスティック回帰のためのデータ

患者 No.	不整脈有無	喫煙有無	飲酒有無	ギャンブル嗜好
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
4	1	1	0	1
5	1	1	0	1
6	1	1	0	0
7	1	0	1	0
8	1	0	0	0
9	0	1	0	1
10	0	1	0	0
11	0	0	1	0
12	0	0	1	0
13	0	0	0	1
14	0	0	0	0
15	0	0	0	0
16	0	0	0	0
17	0	0	0	0
18	0	0	0	0
19	0	0	0	0
20	0	0	0	0

関係性は弱い。

次にロジスティック回帰を行う。表 8 より、ロジスティック回帰を行うと、オッズ比が出力される。オッズ比から順位が把握できるため、不整脈の原因要因の 1 位は喫煙の有無で、次に飲酒の有無となる。ギャンブル嗜好は、不整脈にそれほど影響がないことがわかる。リスク比の結果とオッズ比の結果が異なることは、原因要因が多数ある時は、ロジスティック回帰を使わなければならない。原因要因が相互に無関係と解釈できれば、分割表のオッズ比とロジスティック回帰のオッズ比の順位は同一となる。したがって、要因同士で強い相関がなければロジスティック回帰を適用できる。ここで表 8 の p 値について説明する。p 値が 0.05 以下であれば、今回のサンプル 20 人からでも何十万人という母集団についても、その原因は不整脈に影響を及ぼすと判断される。今回の場合は、どれも 0.05 以上なので、母集団においてこれらが原因であるかどうかはわからない。今回はサンプルが少ないため、サンプル数を増やすことが望まれる。

表 6 多数ある影響要因の比較

(a) 喫煙と不整脈の分割表

喫煙の有無	不整脈		横計	リスク
	あり	なし		
喫煙	6	2	8	75.0%
非喫煙	2	10	12	16.7%
オッズ	3.00	0.2		

(b) 飲酒と不整脈の分割表

飲酒の有無	不整脈		横計	リスク
	あり	なし		
飲酒	4	2	6	66.7%
非飲酒	4	10	14	28.6%
オッズ	1.00	0.2		

(c) ギャンブル好きと不整脈の分割表

ギャンブル好き	不整脈		横計	リスク
	あり	なし		
男性	5	2	7	71.4%
女性	3	10	13	23.1%
オッズ	1.67	0.20		

(d) リスク比とオッズ比

要因	リスク比	オッズ比
喫煙	4.50	15.00
飲酒	2.33	5.00
ギャンブル好き	3.10	8.33

参考文献

- [1] リスク比とオッズ比 https://istat.co.jp/sk_commentary/risk_odds
 [2] ロジスティック回帰 https://istat.co.jp/ta_commentary/logistic

表 7 原因要因相互の分割表

(a) 飲酒と喫煙の分割表

飲酒の有無	喫煙の有無		横計	リスク
	喫煙	非喫煙		
飲酒	3	3	6	50.0%
非飲酒	5	9	14	35.7%
オッズ	0.60	0.33		

(b) ギャンブル好きと喫煙の分割表

ギャンブル好きの有無	喫煙の有無		横計	リスク
	喫煙	非喫煙		
ギャンブル好き	6	1	7	85.7%
ギャンブル嫌い	2	11	13	15.4%
オッズ	3.00	0.09		

(c) ギャンブル好きと飲酒の分割表

ギャンブル好き	飲酒		横計	リスク
	飲酒	非飲酒		
ギャンブル好き	3	4	7	42.9%
ギャンブル嫌い	3	10	13	23.1%
オッズ	0.5	0.40		

(d) リスク比とオッズ比

要因	リスク比	オッズ比
飲酒-喫煙	1.40	1.80
ギャンブル好き-喫煙	5.57	33.00
ギャンブル好き-飲酒	1.86	2.50

表 8 データ例

変数名	回帰係数	オッズ比	有意差判定
喫煙有無	2.643	14.053	0.090
飲酒有無	1.948	7.018	0.162
ギャンブル嗜好	0.504	1.655	0.079