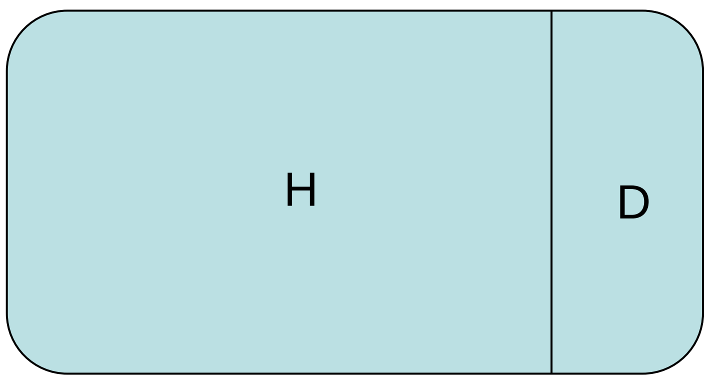
ドクターKの「何でオッズ、どうしてロジスティック？」

　皆さん、オッズ比とかロジスティックモデルと聞いて違和感を覚えることはありませんか。
「オッズ比ってなんだろう、どうしてリスク比じゃないんだろう、オッズとオッズ比は違うのか」などのオッズに関する疑問をよく聞きます。
「ロジスティックって、物流システムのことでしょ。え、数学なの？」もよく聞く質問です。
「数式を最初に持ちだされても納得出来ないよ」というのが本音ではないでしょうか。

今回は、歴史をたどって、皆さんが納得できるように、どうしてオッズ比、ロジスティックを用いるようになったのかをお話します。

オッズとリスクの違い

集団全体



リスク（確率） = D/(H + D)

オッズ = D/H

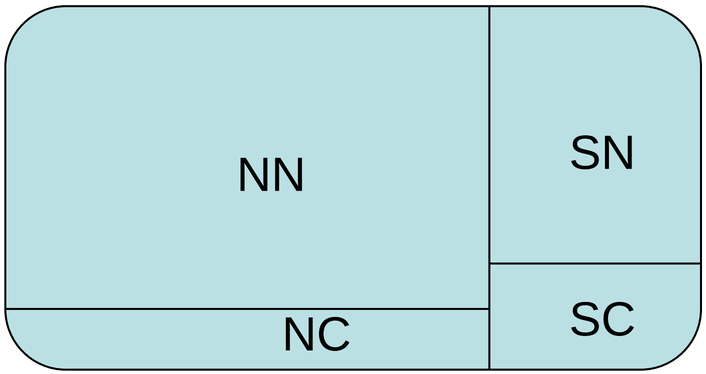
H: 健康な人の割合

D: 病気の人の割合

リスクは病気の人の全体の中での割合、
オッズは病気の人と健康な人の比です

喫煙とがんの2つの分類法を用いると

集団全体



NN: 非喫煙者で非がんの割合

NC: 非喫煙者でがんの割合

SN: 喫煙者で非がんの人の割合

SC: 喫煙者でがんの人の割合

分類法1
喫煙、非喫煙を先に分ける
喫煙者の中のオッズ SC/SN
非喫煙者の中のオッズ NC/NN

分類法2
がん、非がんを先に分ける
がんの中のオッズ SC/NC
非がんの中のオッズ SN/NN

分類法1の場合のオッズ比
(SC/SN)/(NC/NN) = SC NN/(SN NC)

分類法2の場合のオッズ比
(SC/NC)/(SN/NN) = SC NN/(SN NC)

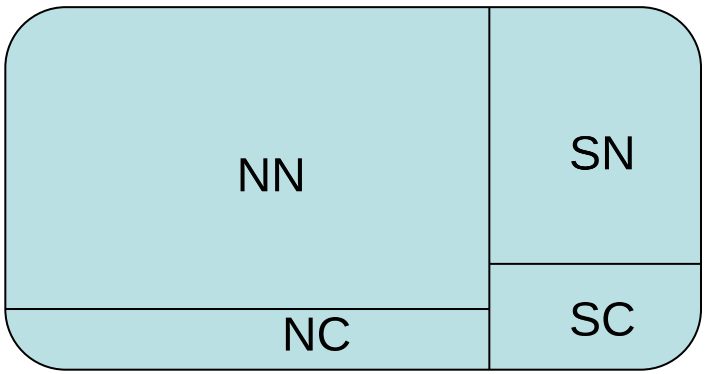
オッズ比の2つの利点

分類法1,2のどちらでもオッズ比は同じ。リスク比は同じではない

症例・対照研究（分類法2）でもオッズ比が求まる

喫煙とがんの2つの分類法を用いると

集団全体



リスク比
(SC/(SN+SC))/(NC/(NN+NC))
= SC/NC (NN+NC)/(SN+SC)

NC << NN, SC << SN の場合は
リスク比 $\simeq$ (SC/NC)/(SN/NN) = オッズ比

NN: 非喫煙者で非がんの割合

NC: 非喫煙者でがんの割合

SN: 喫煙者で非がんの人の割合

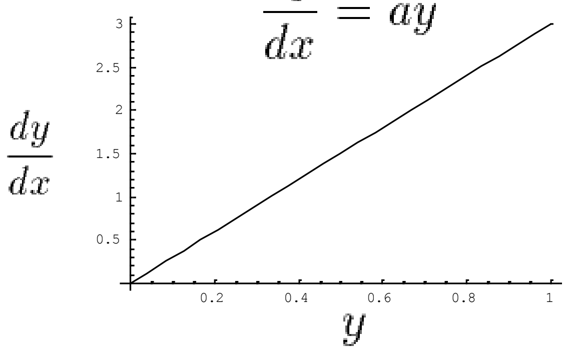
SC: 喫煙者でがんの人の割合

疾患の頻度が非常に低ければオッズ比はリスク比とほぼ同じ

オッズやオッズ比はロジスティックモデルと整合性が良い

マルサスの人口論

$$\frac{dy}{dx} = ay$$



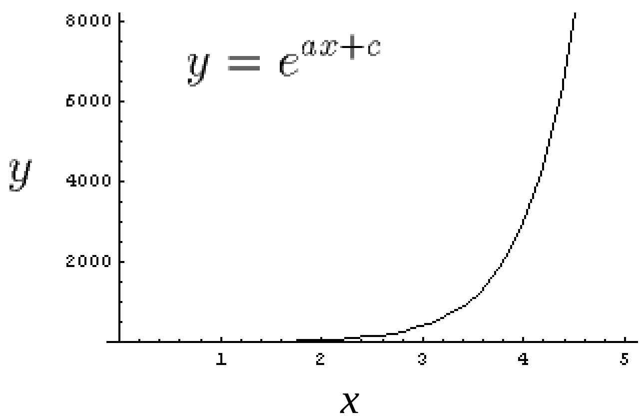
大きさに比例

して増えれば

↓

指数関数になる
（ねずみ算）

$$y = e^{ax+c}$$



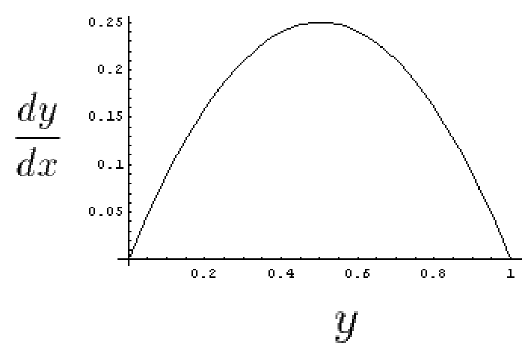
そんなに増えるはずはない

何か制限があるはず。

yがbに近づくと増え方が減少すると言うモデル

↓

$$\frac{dy}{dx} = ay(b-y)$$



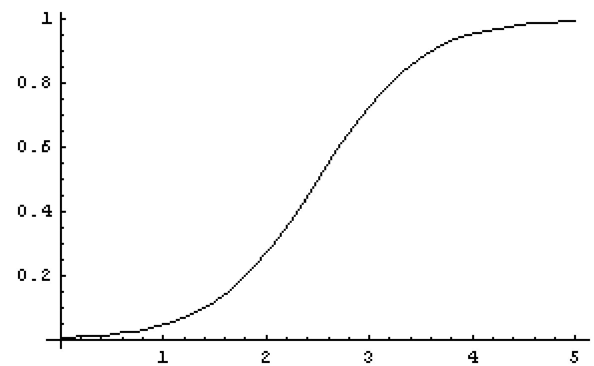
ロジスティック関数

$$\frac{dy}{dx} = ay(b-y)$$

↓ この微分方程式を解くと

$$y = \frac{b}{1+e^{-abx-bc}}$$

ロジスティック関数になる



y が0-1の間だとする（確率のようなもの）と

$b = 1$とすればよいので

ロジスティック関数の逆関数からロジット関数が得られる。
yを確率と考えると、オッズの対数が線形となり便利

ロジット関数

$$y = \frac{1}{1+e^{-ax-c}} \longrightarrow \log\frac{y}{1-y} = -ax - c$$

オッズの対数が線形となる!!!

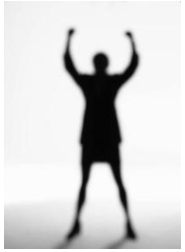
混在した集団から抽出した人はA国人かB国人か？

x cm

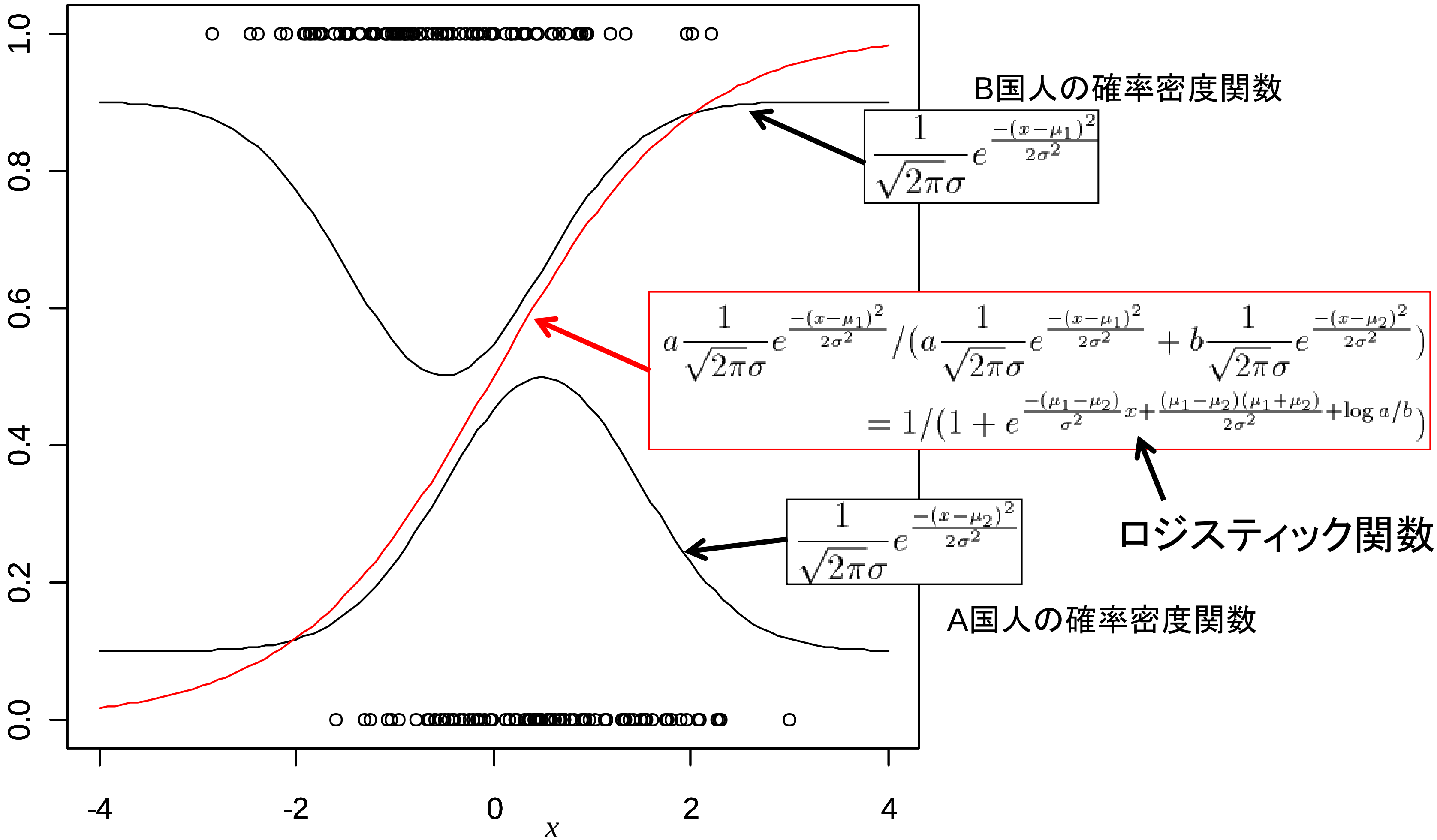
A国人かB国人か不明（身長　172.5 cm）

A国人（身長　173 ± 6.5 cm）

A国人とB国人
が混在



B国人（身長　168 ± 6.5 cm）

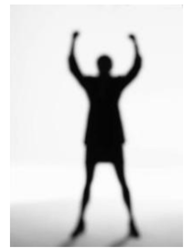


A国人の身長もB国人の身長も平均は違うが分散が同じ正規分布に従うとすると、身長xの人がA国人である確率（事後確率）はロジスティック関数により表される。

a, bは事前確率

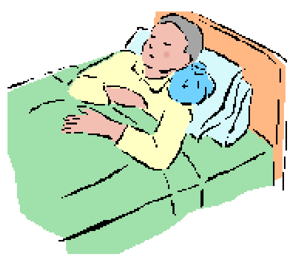
病気になりやすさを示す*Y*と病気との関係は？

*Y*は遺伝型(0,1,2)、環境要因など
の線形結合で、正規分布に従うと
し、*Y*をliabilityと言うことにする。



病気になるか不明(Liability *Y*)

Liability
*Y*の人は病気か？



病気の人

健康な人

健康な人と病気の人のliabilityは平均が違うが分散が同じ正規分布に従うとすると

Liability *Y*の人が病気になる確率は*Y*のロジスティック関数に従う(*Y*は確率のロジット関数)

Liabilityは、対数オッズとなり、オッズはliabilityの指数関数

対数オッズ　Liability　遺伝型などリスク要因

$$\log \frac{\pi_*}{1-\pi_*} = Y = aX + b$$

対数をはずす

オッズ　$$\frac{\pi_*}{1-\pi_*} = e^{Y} = e^{aX+b}$$

Liabilityが相加的だと、
オッズは相乗的になる

オッズ　$$\frac{\pi_*}{1-\pi_*} = e^{Y+\Delta Y} = e^{Y} e^{\Delta Y}$$

$$\frac{\pi_*}{1-\pi_*} = e^{a(X+1)+b} = e^{a} e^{aX+b}$$

Xをリスクアレルの数とすると、リスクアレルが一つ増えるごとにオッズはe^{a}倍になる

対数オッズがリスク要因の線形結合となる

オッズ比の対数もリスク要因の線形結合となる

罹患率が極めて低ければ

オッズ　$$\frac{\pi_*}{1-\pi_*} \simeq \pi_*$$ 浸透率

となり、浸透率が相乗的となる

ドクターKの「何でオッズ、どうしてロジスティック？」

皆さん、これでどうしてリスク（確率）よりオッズ（比）を用いることが多いかわかったでしょう。

オッズやオッズ比の対数がリスク要因の線形結合で表されるという仮定の妥当性もわかったでしょう。

連続ではない2値の形質を対象とする場合、ロジスティックモデルを用いることに十分納得がいけば、次のステップに進むことができます。

納得ができないのに、本に書いてあるからという理由で進むといずれ躓きます。