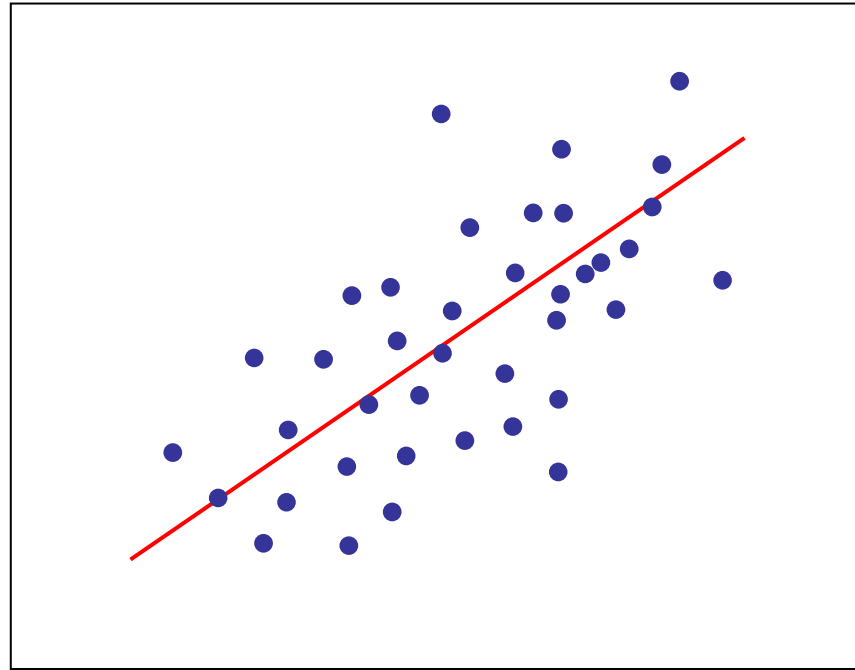


閾値モデルvs. 線形モデル

- 私たちは変数を扱うとき、そのまま連続変数で用いるか(図1)

心血管疾患 線形モデル



収縮期血圧

図1 収縮期血圧と心血管疾患発症リスクの関係
収縮期血圧が上昇するのに比例してリスクが増大している。

- あるいはどこかでカットオフして用いるのか(図2)、

心血管疾患

閾値モデル

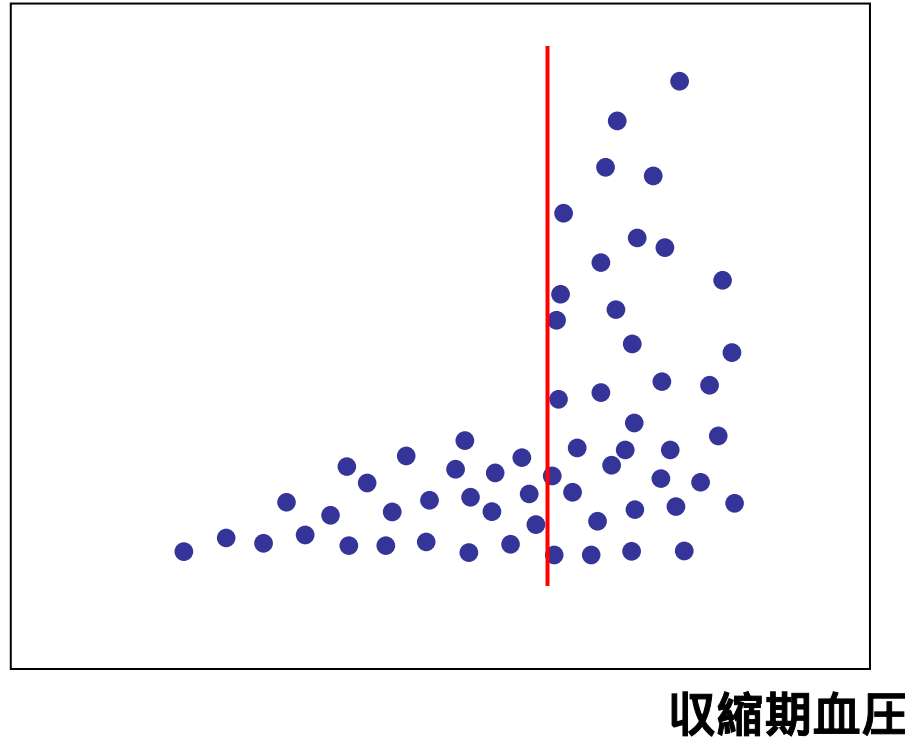
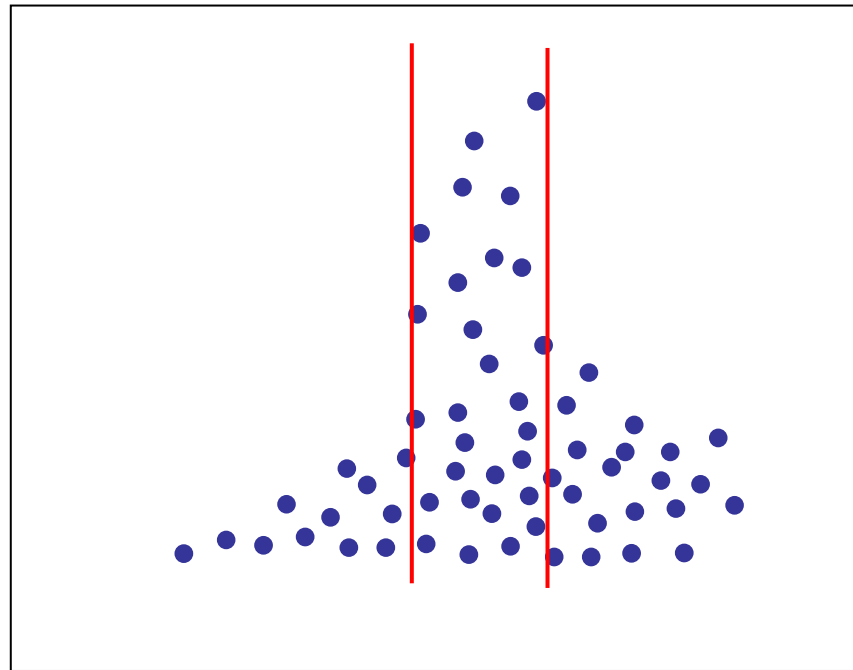


図2 収縮期血圧と心血管疾患発症リスクの関係
収縮期血圧が一定の閾値を越えるとリスクが急増している。

- あるいはいくつかに分けてトレンドの有無(図3)をみるのか決断しなくてはならない。どれがよいということはなく、ケース・バイ・ケースで比較してみる必要がある。

女性化現象

特殊



内分泌攪乱物質

図3 内分泌攪乱物質濃度と男性の女性化現象リスクとの関係
内分泌攪乱物質濃度が中等度るとき女性化現象のリスクが
むしろ高くなる。濃度が高いからといって、女性化現象が強
くなるとは限らない。

- 一般的に高血圧は140mmHgを超えると治療の対象となる。WHOと国際高血圧学会は130mmHg以下を正常とし120mmHgが適当としている。この基準は血圧が上昇することによって心血管疾患が増えることに基づいている。さらにこれらの方針はフラミンガム研究で得られた知見に基づいている。このデータをもう一度紐解き、血圧と心血管疾患の関係が線形モデルなのか閾値モデルなのかを検討したいと思う。Portらは線形モデルより閾値モデルが適切だと主張しているが(Systolic blood pressure and mortality. Lancet 2000 15;355:175-80)、本当だろうか？心血管疾患を発生した1000人に対して2000人のコントロールが選ばれた。

線形モデル

- 線形モデルとは、図1のように血圧が上がれば上がるほど、心血管疾患のリスクもあがるようなものをいう。もちろん、血圧の実測値を連続変数として扱えばsimple linear regression となり、直接的である。しかし、今回はWHOの基準を参考に以下のカテゴリー(indicator variable)に変換したいと思う。
- <120 mmHg, $120 - 129$ mmHg, $130 - 139$ mmHg, ≥ 140 mmHg
- そして、収縮期血圧が120 mmHg 未満を基準に他の血圧カテゴリーでリスクがどのように変化するか検討してみた(表1)。

表1 . 心血管疾患発症と収縮期血圧 (WHO基準) との関係 (線形モデル)
 Log likelihood = -1822.8823

収縮期血圧 (mm Hg)	coefficient	95% CI
<120 mmHg	0	
120 – 129 mmHg	0.29	0.02 ~ 0.56
130 – 139 mmHg	0.87	0.62 ~ 1.12
140 mmHg	1.43	1.17 ~ 1.69
cons	-1.40	-1.62 ~ -1.19

- 血圧が上がれば上がる程、心血管疾患のリスクも上がる (coefficient of the 95% CIは0を含むか否かで有意かどうか判る)。しかし、140 mmHg で心血管疾患の発症リスク上昇が顕著である。実測値をカテゴリーに分ける場合、それは研究者の裁量である。血圧120 mmHg 未満のグループは他よりどうしても少なくなってしまうので、研究者によってはパーセントイルで4等分してみるかもしれない。同様の傾向をしめしたが、WHOの基準に従ったモデルの方が優れていた。
- 次に<120 mmHg, 120 – 129 mmHg, 130 – 139 mmHg, >=140 mmHg を、それぞれ0, 1, 2, 3 というカテゴリーに置き換え解析してみた (表2)。有意差があり、表1にモデルとして同等であった。一方、「0, 1, 2, 3」ではなく、それぞれのカテゴリーの平均値を代用することもできるかもしれない。これも「血圧が上がれば上がるほど、心血管疾患のリスクが上がる」ことを示している。

表 2 . 心血管疾患発症と収縮期血圧カテゴリーとの関係 (トレンドモデル)
Log likelihood = -1824.4476

収縮期血圧 (mm Hg)	coefficient	95% CI
血圧カテゴリー	0.51	0.43 ~ 0.59
cons	-1.54	-1.70 ~ -1.38

閾値モデル

- それでは、血圧140 mmHg で区切った閾値モデルではどうだろう。最初のモデルと変わらないどころか有意に劣化してしまった(log likelihood ratioが0に近いモデルの方がよい。逆に0から遠ざかるものは悪い。モデル間の比較は、 χ^2 で行う)。よって閾値モデルよりは線形モデル(トレンドモデル)を採用すべきだろう。

表3 . 心血管疾患発症と収縮期血圧との関係 (閾値モデル)

Log likelihood = -1851.7771

収縮期血圧 (mm Hg)	coefficient	95% CI
140 mmHg	0.96	0.79 ~ 1.14
cons	-0.93	-1.02 ~ -0.84

年齢という交絡因子

- 収縮期血圧と心血管疾患発生リスクの間で年齢は交絡因子である可能性はないだろうか？つまり年齢と収縮期血圧は正の関係にあり、高齢は心血管疾患発生に寄与するという構図である。年齢を調整するには、先のモデルに年齢の変数を加えるだけでよい(表4、表5、表6)。

表 4 . 心血管疾患発症と収縮期血圧との関係 (年齢で補正)

Log likelihood = -1686.8102

収縮期血圧 (mm Hg)	coefficient	95% CI
年齢	0.07	0.06 ~ 0.08
<120 mmHg	0	
120 ~ 129 mmHg	0.29	0.02 ~ 0.55
130 ~ 139 mmHg	0.67	0.40 ~ 0.93
140 mmHg	0.89	0.61 ~ 1.16
cons	-5.22	-5.76 ~ -4.68

表5 . 心血管疾患発症と収縮期血圧カテゴリーとの関係 (トレンドモデル)
Log likelihood = -1687.1579

収縮期血圧 (mm Hg)	coefficient	95% CI
年齢	0.07	0.06 ~ 0.08
血圧カテゴリー	0.30	0.22 ~ 0.39
cons	-5.20	-5.70 ~ -4.70

表6 . 心血管疾患発症と収縮期血圧との関係 (閾値モデル)

Log likelihood = -1700.6877

収縮期血圧 (mm Hg)	coefficient	95% CI
年齢	0.07	0.06 ~ 0.08
140 mmHg	0.48	0.29 ~ 0.67
cons	-4.98	-5.48 ~ -4.48

- ロジスティック解析におけるcoefficient は単純にオッズ比(OR)に変換できる。
- 例えば、表2のcoefficient = 0.51 は $e^{0.51} = 1.66 = \text{OR}$, 表2のcoefficient = 0.30
- は $e^{0.30} = 1.34 = \text{OR}$ と置き換えられる。この場合、年齢で補正することによりORは1.25倍変化しているから、年齢は交絡因子であると考えられる。

- 以上の行なってきた解析結果から考察するかぎり、血圧が140mmHgを超えると急に心血管疾患のリスクが上昇するわけではなく、血圧が高くなればなる程心血管疾患リスクもあがると考えた方が良さそうである。数値を連続変数あるいはリネアトレンドとして捉えるのか、カテゴリーでとらえるのかで随分結論が異なることがある。よって両方で検討すべきである。カテゴリーの場合はどこでcut-offを引くかでコントラストが違ってくるが、散布図を眺めてカットオフ値を決めるなどする必要があるかもしれない。論文を読んでいるだけでは、どのような方法を試したのかを知ることとはできない。しかし、いろいろな角度から解析してみることは重要で、それまでのデータを集めるまでの苦勞と比べればたいしたことはない。統計解析依頼者は簡単に考えているが、解析には時間がかかるものである。