

最近のポスターから...

そもそもがんの発症に関しては

- ・家族や〇親等以内にがんの人がいる(遺伝的)
- ・〇〇を食べるとがんになる(環境的)
- ・〇〇な人がなりやすい(恣意的?)

↓↓

などなど色々なことが言われておりますが...?

それを明確に**証明できる方法がありません!**それゆえ、がんの発症や治癒に関しては、**おかしな主張が入り込む余地が非常に多くなってきます**

がんが目をつけられやすいのは、非常に重大な病気かつ身近な存在(男性の60%、女性の40%が一生のうち罹患する)であるからに他なりません  
⇒サギの人々にとっては不安の煽りやすさに他ならない!

基本的には**遺伝+環境**要因

シユレルンによる**輻輳説**に例えるなら...

がんは遺伝的要因と環境的要因(食べ物や生活様式など)が、ちょうど半々に作用するようところにあり、その原因については説明が難しい

地域・職場などある共通の集団でまとめて発症するといった現象が発生しない限りは、たった一つの「**何か**」に原因を求めることは不可能です!

例)こんな人ががんになりやすい

- ・枕元に金魚鉢が置いてある
- ・高級車に乗っている
- ・〇〇市に住んでいる

⇒(マスコミに乗っかりやすいネタでもあります)

実はこのようなからくりが...

- ・入院患者が枕元に金魚鉢を置いただけ
- ・金持ちゆえに検診を受ける機会が多い
- ・〇〇市には終末期ケアセンターが存在する

不幸にしてがん罹患してしまった人々の特徴を、後付けで**何でもい**から集めただけです。しかし**数字上はウソ**ではないものも多く、それゆえに**厄介**だったりします

研究の観点でも、がんの発症と「何か」が間違いなく結びつけば、その「何か」は極めて重大な発見になります。それゆえ、研究者は自分が目をつけた「**何か**」と**強引に結び付けよう**とします

**本日はそんな事例を見ながら、統計の正しい利用について学びましょう!**

〇〇人には、△△が多いんです!  
そんなの誰が決めたんですか?

© 2007- 2012 Nippon BS Broadcasting Corporation. All Rights Reserved.

勘違いされやすい・・・相関

x、yの相互関係の強さを見る・・・もの、それを表すのが「相関係数:r」です

数字上の関連の強さなどいくらでも持ってくる事が出来ます

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$

有名なPearsonの積率相関係数です。必ず-1 ≤ r ≤ 1となり、|r|が1に近いほど「相関が強い」ということです

具体的にいくつ以上なら相関がある、強い、ない、弱いなどの基準はありません。一応の目安としては・・・

|r| < 0.3で「相関は見られない」、0.3 ≤ |r| < 0.5ぐらいで「若干の相関あり」、0.5 ≤ |r| < 0.7で「相関が見られる」、|r| ≥ 0.7で「強い相関が見られる」

ということが多いようですね

その有意性(ここまで見ていないパターン多し)

母集団の相関係数をρとした場合、帰無仮説H<sub>0</sub>: ρ=0、対立仮説H<sub>1</sub>: ρ ≠ 0を検定する。検定統計量は、

$$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

により求められる。これは自由度(n-2)のt分布に従うので、  
|T| > t<sub>n-2</sub>(α/2) (=両側検定)であるときに、有意水準αで帰無仮説を棄却する

求めた相関係数は使い物になりますでしょうか(=有意でしょうか)？

基本的にrが大きくなればなるほど、nが多くなればなるほど有意になりやすいですが、多くの皆様がこちらに関しては無頓着です。異常なほど少ないサンプル数で求めた相関はあてになりません

p>0.05のときには、どれほどr値が大きくてもxとyの関連が見られないことになります！

これは何を検定するのか？

$$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

r: サンプルから求められた相関係数  
n: 相関を求めるのに用いたペアの数(サンプル数)

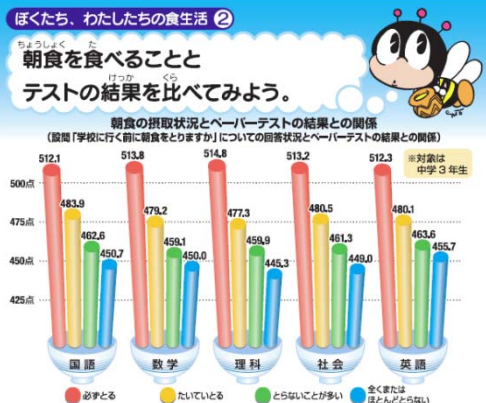
帰無仮説(H<sub>0</sub>): 母集団の相関係数はゼロである

対立仮説(H<sub>1</sub>): 母集団の相関係数はゼロでない

上記の式で考えますと、rが1もしくは-1に近いほど、nが大きければ大きいほどTの値は大きくなります。

早い話が、サンプルが大勢いれば有意にはなるということです。t検定のようにn数に依存しますので、統計的有意=臨床的有意ではないことに、改めてご注意願います

文部科学省のキャンペーン

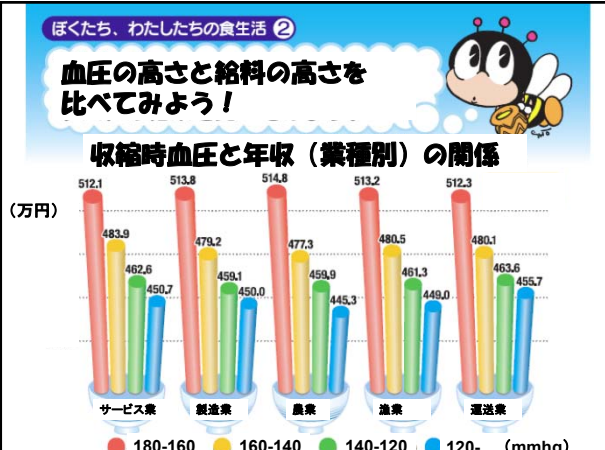


よし、これからもしっかりと朝食を食べよう。

朝食は食べて成績アップだね！

朝食をきちんととっているほうが、各教科ともテストの結果が良いことがわかります。

すこ はやおき 少しでも早起きをして、きちんと朝食を食べるようにしよう。



よし、これからはお父さんの血圧を上げよう！

非行全書 血圧上げて、年取アップだね！

年取が高い人の方が、どの業種も血圧が高い傾向にあることがわかります

**料理の塩分を増加させたり、怒らせるなりして、どんどんお父さんの血圧を上げよう！**

**相関係数が高い&p値<0.05=因果関係を示すではない！**

- 肥満者はアニメを好む
- ガムを噛むと不良になる
- 血圧が高いと給料が高い
- 金魚を飼育するとがんになる
- AHの曲を聴くとバカになる
- 巨人が勝つと景気が良くなる
- 金持ちはがんになりやすい

ならば力士は皆アニメファンか？野球選手は皆不良か？ならばみんな血圧上げるか？入院と共にベッドサイドに置いていただけ？そもそもバカの定義が不可能？カネにモノを言わず球団だから、契機の良いときか勝てないのでは？がん検診のお金があるから、発見しやすいのでは？

**事象間の結びつきは強く、有意か？**  
時間的には原因が先行しているか  
それらの関係はいつでもどこでも変わらないか(普遍性)  
そもそも合理的と言えるだろうか(合理性)

**全てを満たしていても、実は最後にハネられることが多い！**

**因果関係(どちらが先か?)**  
Amazonではおなじみの光景なのですが...

よく一緒に購入されている商品

この本と同時購入した統計学入門—SAS Enterprise Guideユーザーズガイド 高橋 真央 著 ¥ 4,920 をあわせて買う

合計価格: ¥ 9,985

商品カードに入れる

この商品の買った人はこんな商品も買っています

ページ 17 / 17 商品

読者群別データの統計科学—因果推論・機械学習—データ分析の科学 黒野 孝志 著 ¥ 3,990

プロフェッショナルの真実—いかに成果をあげ、成長するか 日比野 健一 著 ¥ 1,890

その数学が理解も決める(文庫文庫) イアン・アーツ 著 ¥ 770

わかる数学 三章 藤原 俊 著 ¥ 1,680

著者: TAO HEARTS VOICES - 稲垣 謙一 著 ¥ 2,800

私の本を購入された方のうち1人が購入した商品がリストアップされた結果。しかし、稲垣 謙一さんのCDを購入された方の多くが、私の本を購入しているわけではない

**如何にも尤もらしいが？**

朝食をきちんと食べ、規則正しい生活をさせる家庭は教育環境に配慮する可能性が高く、それが成績の高さに影響した  
⇒良好な教育環境への配慮が「朝食」と「成績向上」の双方に影響したと考えるのが妥当

「成績向上」は「朝食」によるものであるという解釈は、いわゆる「偽相関」。朝食即成績向上のような話ではなく、同様に「〇〇な人ががんになりやすい」など、**単純な理由では片付かない！**

血圧の事例ぐらいおかしければ納得して頂けるでしょうけれども、このお話のように大々的なキャンペーンが打たれ、しかも**政府が主導している**ような場合ですと、それらしく見えてしまうのが怖いところ  
⇒実はこのメカニズムは、論文に通じるところでもあります

**論文の読み方 (Critical Readingのススメ)**

テレビ大賞 <http://www.tv-osaka.co.jp> © 2009-2017 Television Osaka / Television Sakamoto

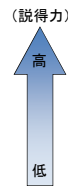
**そもそも論文とは？**

- 事実を積み重ねて結論を導き出す行為が研究であり、その結果をまとめたものが論文
- 「私はこう思う」などの主観を綴っているだけでは、既にScienceではなくなるので、論文ではない
- 仮に何か主張をしたいのであれば、客観的な結果を積み上げて示すしかない
- 数字で語られている=客観的な結果ではない。数字そのものよりも、その数字がどのように集められ、どのように解析されて求められた数字であるかの方が大切
- 時に自分に都合の良い数字を作り上げて(でっち上げてではない)、客観的な結果に見せてしまう方法はいくらでもある

「ありがたや」になっていないか？

- 論文を書いているということは、それだけでも凄いことだ。読むだけでも精一杯の自分から見れば、吟味なんて生意気なこととは言えない
- 大学の先生が書いているのだから、そんないい加減なことは言わないだろう
- 論文には厳しい審査があるはずだから、そんないい加減な結果が独り歩きすることは無いだろう
- 書くだけであれば決してすくは無い。むしろ、日本語論文などは投稿がなくて困っている雑誌もあるのだから、どんな内容であれ掲載する雑誌はいくらでもある
- 立場は関係なくいい加減なことを言う場合はある(それが意図的であるか無いかは別)
- 審査は決して厳しくない。無いものさえもある

ならば・・・Evidence Level



レベル	研究のデザイン
I a	Randomized Clinical Trial (RCT) のメタアナリシス
I b	RCT (ランダム化比較試験)
II a	デザインされた比較研究 (非ランダム化)
II b	デザインされた準実験的研究
III	デザインされた非実験的記述研究 (比較研究・相関・症例研究等)
IV	専門家・権威者の意見・経験

(Oxford Centre for Evidence-based Medicineの分類(2001)より作成)

メタアナリシスについては「後付け」ということで、本当にEvidence Levelが一番高いのかと言えば、そうでないという説もあります。が、一番注目すべきは、**専門家や権威者の意見が一番下**になっている点ではないでしょうか？

論文を読む姿勢(Critical Readingのススメ)

早い話が、きちんとデザインされた研究を行えば、専門家や権威者の意見や経験以上のEvidenceとして社会から受け入れられるということです

それがきちんと守られているか。皆様はそれなりの視点を持って論文を読まなければなりません

まずは論文の構造に関する掟(?)となる、**IMRAD構造**について学びましょう。IMRADとは、**Introduction(緒言)**、**Method(方法)**、**Result(結果)** And **Discussion(考察)**の略です

特に何が大切というのはありませんが、各パートにきちんと下記の情報が記載されているのかを確認しましょう！

Introduction(緒言)

- 研究を行うに至った経緯や背景などが記載されているか  
⇒そもそも筆者はなぜこの研究に興味を持ったのか、問題意識は何なのか？
  - この研究の目的や仮説、新規性などは記載されているか  
⇒何を調べて、何を明らかにして、何が疑問で、何よりも、この研究は従来のモノと比較して何が新しいのか？
- ここがつまらないと、Reviewerにも読んでもらえません。Rejectになる論文はここが不明瞭であり、ほとんどのものでは新規性が感じられません  
⇒**新規性に関しては、事前のリサーチが非常に大切です。先行研究などは調べてから行わないと「同じものがあります」という指摘を受けてしまいます**

Method(方法)

- 研究の対象は何か
- 研究の対象となった集団についてのプロフィール(年齢、性別、既往歴など)が明確に記載されているか
- その集団が選択された理由および、別の集団が選択されていない理由は何か
- 主要評価項目、副次評価項目はそれぞれ何であり、仮説や問題意識とはリンクしているか
- 統計処理を行う場合はその手法が適切なものであるか  
⇒例)本来は生存時間解析を行うべきところを、単なるχ2検定などで済まされていないか？
- 介入・観察・割付の方法はどうか(ランダム化、盲献化などを行っている場合は、その方法や遵守の度合いなどを示しているか)
- 利用したデータの出所は明確であり、信頼できるものか

Result(結果)

- 仮説や研究者の疑問は解決できているか
- IntroductionやMethodにおいて記載されている疑問と対応しているか
- 明らかな論理の飛躍や、結果に対する強引な意味づけはなされていないか  
⇒たとえば、 $p=0.051...$ 付近で「有意ではないが傾向が見られる」などの記述をすること
- 解析対象となった集団の数は明確に記載されているか



(参考) Analysis (統計処理)

- 特定の集団に対して有効であることが主張されている場合は、そうでない集団に関して言及がなされているか(層別解析)  
⇒たとえば、「30歳未満の女性には非常に効果が高かった」なる結論の場合は、**結果の一般化が困難**になる
- 後付の解析になっていないか  
⇒当初の予定にはなかった解析を行っているか?
- アスタリスク以外に、p値は記載されているか
- 検定の多重性は評価されているか  
⇒探索試験であればどれほどの検定を繰り返しても「参考程度」なので問題はない。検証型試験において、いくつも評価項目があったりしないか

**掲載されているp値はどのように導かれたものかは、読む人がそれぞれ理解することです!**

Discussion (考察)

- 他の研究や先行研究と比較して、当該研究による新しい知見は何だったのかが明確に記載されているか
- 率直に、感心・共感できたか
- 結果および新しい知見の重要性について記載されているか**

こうして見ると、Methodのチェックポイントがかなり厳しいですね。つまり、方法論がきちんとしていない研究は、結果がきちんとするはずが無いということです!

皆様の役割は...実は大きい

論文を読む、書くなどは自分には関係ない...のではない!  
部外者であることは絶対にやめよう!

皆様が論文を見る目を養うということは...?



いい加減な論文が認められにくくなるということ。つまり...  
いい加減なもの存在を認めなくなる人々が多くなるということ。  
つまり...  
それだけ医学の発展に貢献するということです!

たこまる

**生物統計学とは?  
(統計的検定<<デザイン!)**

テレビ大阪 <http://www.tv-osaka.co.jp> © 2009-2015 Television Osaka, Inc.

医学研究とは?

臨床研究・疫学研究などが総称として呼ばれるようになるようです

- 「臨床研究」とは、人を対象として疾病の予防方法、診断方法や治療方法の改善、疾病原因や病態の理解、患者のQOL向上を目的として実施される医学研究
- 「臨床研究」には人由来の細胞や組織等を使った研究も含まれるが、特に「人の集団」を用いて行う場合を特に「**臨床疫学研究**」と呼ぶ

例:私の好きな遺伝子解析などは人を対象としておりますが、扱っているのは「人の組織」ですので、「臨床研究」になります。流行のiPS細胞などもそうです

臨床研究・疫学研究・臨床試験・治験?

- 臨床疫学研究のうち、**対象集団を複数の群に分け**、それぞれの群に異なった予防法、診断、治療等を与え、比較する研究のことを「臨床試験」と呼ぶ
- 臨床試験のうち、薬剤や医療機器として**承認されるために**有効性や安全性を確かめるためのものが「治験」



一番条件が厳しいのは治験です。いわゆるGCPなどにより、厳重に規定されております

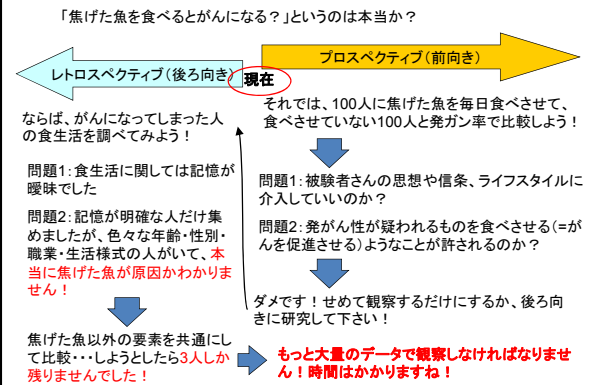
どうやら、「医学研究」というのは、これらの総称になるようですが、本来は「臨床研究」ということなのでしょうが?

(臨床研究人材教育コンソーシアムのHPより参照)

### 臨床研究にはどんな方法がある？

- ◆過去のデータから何かを探索する？  
(後ろ向き研究:レトロスペクティブ研究ともいう)  
・データは全部用いる？  
・必要な条件に従って抽出する？
  - ◆これから新しく研究を行って探索・検証する？  
(前向き研究:プロスペクティブ研究ともいう)  
・とにかくやってきたデータを観察するだけ？  
・介入する？  
→介入する場合、患者さんのライフスタイルに影響を与えてしまうから、倫理上の問題を強く問われます！
- とにかく目的に合致したデザインが必要！特に都合の良い対象だけを抽出する(=Bias)には要注意です！**

### レトロとプロのイメージ



### 横断研究(ある1時点のデータを見てモノを言う)

どうやら、レトロもプロも向きがあるということらしい

↓

おなじみのt検定やχ2検定などは、もしも何の解説もなければ「どの時点のものか定かではない」ということになる！

↓

つまり、それはある時点に関する話だけであり、その過去や未来のことは、第三者にはわからない！

もちろん、研究目的が「その時点のこと」を語るということであればOKですが、時間的経過などを見たい場合はこのデザインはダメ！

### こんな事例

- ◆何とか有意差を出してくれないと困る！
- ◆どこかで有意差が出てくれるような群分けはないものか？
- ◆この群にだけ限れば有意差は見える！
- ◆どのデータを除外すれば有意差が出てくれそうか？
- ◆有意差が出やすい検定方法はないか？
- ◆有意差が出ないのはお前のせいだ！
- ◆何とか症例数が1000症例になるような証拠を見つけてくれ！(後述)

後付け、Ad-hoc、泥縄、こじつけ…色々な表現が出来ますが…

**ほぼ全てが、適切な研究デザインを怠った結果です！**

### そもそも研究の目的は？

研究にはいくつかの目的があり…まずは知りたいことがあるはず？

↓ ⇒(それがなければここで終了！)

- ①何らかの仮説を持っていて、それを明らかにしたい  
⇒ならば、**比較対照群を設定した介入型の臨床試験**をしなければなりません。意味の無い差に反応しないように、仮説や呼び調査に基づいて、症例数を設定して行ないましょう！
- ②現段階で仮説はありませんので、何が大事か探索したい  
⇒ならば、**評価項目を多めに設定した探索型の試験(観察)**を設定しましょう。ただし、介入型のような説得力もありませんし、**p値がいくつ有意になってもそれは参考程度です！**

### Keyword 「手当たり次第」

インチキな祈禱師はがん患者に近づく(どうやって?)

がんセンター病院などの駐車場で車のナンバーを調べ、持ち主を割り出して**手当たり次第**に電話をまくる。反応があった人間をカモにする。家族を救いたいという気持ちも手伝っているので、引っ掛かりやすいようだ  
また、神社の絵馬に住所や氏名、電話番号や病名まで書いてあるものがある。それらを全てひかえて連絡しまくる

(家族)どうして家族ががんとわかったのですか？  
(祈禱師)私の超能力で全てお見通しです。等教団の〇〇を購入すれば…



もしもそのうちの一人でも、例えば手術や化学療法、もしくは原因不明でも回復した場合には「私の祈りが通じた」と主張すればよい。それがインチキであることの証明方法はない！

**手当たり次第に検定しろ！**

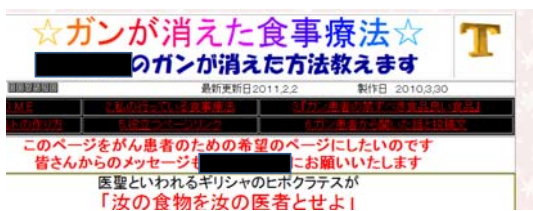
- ヨーロッパの西の片田舎の貧家に男の子が生まれる。彼は弁舌で多くの人の心をつかえ、その名声は東の王国にまで鳴り響く(ノストラダムスはこんな予言が3000種類?)
- 11Rマイルチャンピオンシップ 京都競馬場 18頭立て ある予想屋
  - ⇒「こっそり教えてあげよう。明日の11Rの馬連は1番が勝つよ」・・・これを18人の人に別々に伝達したら?
- 「明日は雨が降るよ」(翌日晴れ)・・・「世界のどこでとは言っていない！」
- 「私は超能力者だ。私がお祈りすれば必ず雨を降らせることができる」・・・雨が降るまで祈り続けられればよい。雨が降ったら「私の祈りが通じた！」と言えば良い

**統計的検定の価値とは？**

- 比較対象群がA~Jの10群もある臨床試験  $_{10}C_2=45$ 回のt検定で「どこかで有意差が出ました」
- 評価項目 (Endpoint) が10項目もある臨床試験  $_{10}C_2=45$ 回のMann-WhitneyのU検定で「どこかで有意差が出ました」
- 評価時点が5時点の臨床試験  $_{5}C_2=10$ 回の対応のあるt検定で「どこかで有意差が出ました」
- 20例でt検定を行って有意にならなかったのに、もう1000例追加してやったら有意になりました
- 化学療法Aと科学療法Bについて、5年後の生存率を比較するために、年に1回ずつ×2検定で比較して、1年目、2年目は有意でしたが、3年目以降は有意ではありませんでした

(参考) 例えばこのようなWebsite

誰が見ても胡散臭そうに見えるわけですが...



この場合は手当たり次第に見せて、反応のあった人間をTargetにするわけです。きっかけはともかく、もしも治癒した人がいたら、その人を広告塔にすればよい

某国立大学病院 試験実施計画書ひな型より抜粋

評価項目

有効性等に関して主要評価項目、副次的評価項目を記載します。また、安全性評価項目を記載します。  
**しばしば見られるのは、できるだけたくさんデータをとって統計解析を繰り返して、有意差が出たデータで論じようとする場合です。しかし、何度も検定を繰り返すと本当は差がないのに偶然有意差が出る可能性が高くなります(検定の多重性の問題)。**  
 また、仮説を立てる段階で、説明変数(原因と見なす項目、薬物治療など)と明らかにしたいアウトカム(結果と見なす項目、心血管イベント発生率、QOLなど)を設定してははずです。明らかにしたいアウトカムを主要評価項目として設定する必要があります。その他のアウトカムを副次的評価項目として記載します。設定根拠の記載も望めます。



そんなに「しばしば見られる」現象なのか?  
 (参考までに、ウチではありません)

**Endpointの種類(Hard、Soft?)**

評価項目にも色々あります。たとえば・・・

- |  |   |   |
|--|---|---|
| <ul style="list-style-type: none"> <li>➢5年後の生存率</li> <li>➢3ヶ月後の体重減少率10%以上</li> <li>➢脳出血の有無</li> </ul>  | ➡ | これらの指標は誰が見ても明確なもの (HARD Endpoint)                     |
| <ul style="list-style-type: none"> <li>➢入院の有無(入院を決定するのは医師であり、医師の判断が入り込む余地アリ)</li> <li>➢目視による傷の改善度が10%程度以上</li> <li>➢「楽になった」患者の割合が10%以上の増加</li> </ul> | ➡ | これらの指標は極めて曖昧で(個人の判断が入る、または明確なモノサシがない) (Soft Endpoint) |



### 生存時間解析 そもそも・・・

生存「率」なのだから、何故  $\chi^2$  検定ではいけないのか？  
 5年後に「生存」「死亡」を、2群間で比較すればよいのでは？

	生存	死亡	合計
治療法A	60	40	100
治療法B	30	70	100
	90	110	200

X2検定  
p<0.0001

- ・大切なのは5年後の結果だけではない。5年後の単純比較では、治療の翌日死亡したのも4年と364日で死亡したのも**評価は一緒**である
- ・A群は1年以内にほとんどが死に、B群は4年経過以降にほとんどが死んだかも知れない
- ・死亡以外の打ち切りを評価していない。分母はどれだけ脱陸したのか？

### 生存時間解析(1)

ココで言うところの「生きている」という意味？

別に文字通りの「生死」に限った話ばかりではない。その出来事をもって「観察終了」となる出来事が発生するまでの時間を「生存時間」と呼んでいる

◆観察の終着点ということで「Endpoint」ともいう。



通常、Endpointとは「評価項目」を示す。最も重要な評価項目を「Primary Endpoint(主要評価項目)」、その他の評価項目を「Secondary Endpoint(副次評価項目)」と呼ぶ  
 (これによって検定方法や必要な症例数が決定されます)

### 生存時間解析(2)

生存時間解析→そもそも生存時間とは何か？

基準となる時点からある事象(イベント)が発生するまでの時間の長さです。例えば・・・

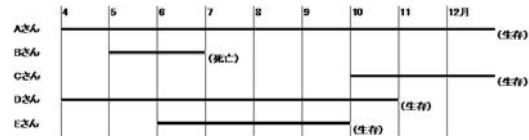
- ◆手術から死亡までの日数
- ◆薬剤の投与から治療効果が認められるまでの時間
- ◆入門から入幕までの場所数



生存時間という呼び方はしておりますが、基本的に生死のみを扱うわけではありません。ある時点から、観察すると決めた事象・出来事(=イベント)が発生するまでの時間のことを、便宜上「生存時間」と呼んでいるのです。

### 生存時間解析(3)

5人の肺がん患者を観察した結果、下記のようにになりました



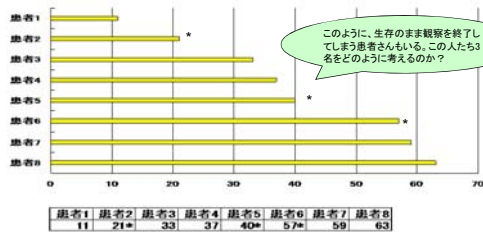
- ◆Aさん：観察開始から観察終了まで生存(9ヶ月)
- ◆Bさん：5月に入院、7月に亡くなられました(2ヶ月)
- ◆Cさん：10月から観察終了まで生存(3ヶ月)
- ◆Dさん：観察開始から10月で転院されました(7ヶ月打ち切り)
- ◆Eさん：観察開始から9月で引越されました(4ヶ月打ち切り)



実際に様々な理由で観察できなくなります。さらに観察スタート日もまちまちですし・・・？実際に例題で理解することにしてしまおう！

### 生存時間解析(4)

8人の肺がん患者を観察した結果、下記のようにになりました



基本的には、大切なのは「長さ」です。いつから観測しているかではなく、生存時間(例：何日生きていたか、何日後に発生したか、何場所で昇進したか)が重要なのです！

### 生存時間解析(5)

患者1	患者2	患者3	患者4	患者5	患者6	患者7	患者8
11	21*	33	37	40*	57*	59	63

\* 生存のまま観察打ち切り

日付	T(0)	T(1)	T(2)	T(3)	T(4)	T(5)
T(n)における死亡数	0	1	3	3	5	6
T(n)直前の生存数	8	8	6	5	2	1
T(n)からT(n+1)における打ち切り数	0	1	0	2	0	

- ・ $0 \leq t < 11$  (=患者1の死亡直前まで) 区間の生存確率は、 $(8-0)/8=1$ となる
- ・ $11 \leq t < 33$  (=患者3の死亡直前まで) 区間の生存確率は、 $(8-1)/8=0.875$ 。よって、この時点までの生存確率は  $1 \times 0.875=0.875$ となる
- ・ $33 \leq t < 37$  (=患者4の死亡直前まで) 区間の生存確率は、 $(6-1)/6=0.833$ 。よって、この時点までの生存確率は  $1 \times 0.875 \times 0.833=0.729$ となる
- ・ $37 \leq t < 59$  (=患者7の死亡直前まで) 区間の生存確率は、 $(5-1)/5=0.8$ 。よって、この時点までの生存確率は  $1 \times 0.875 \times 0.833 \times 0.8=0.583$ となる
- ・ $59 \leq t < 63$  (=患者8の死亡直前まで) 区間の生存確率は、 $(2-1)/2=0.5$ 。よって、この時点までの生存確率は  $1 \times 0.875 \times 0.833 \times 0.8 \times 0.5=0.292$ となる
- ・ $63 \leq t$  (=患者8の死亡直後以降)の生存確率は、 $(1-1)/1=0$ 。よって、この時点までの生存確率は  $1 \times 0.875 \times 0.833 \times 0.8 \times 0.5 \times 0=0$ となる

イベント発生以外の打ち切りを考慮しないと結果が矛盾します！



生存時間解析(6) たたとえばこんな事例

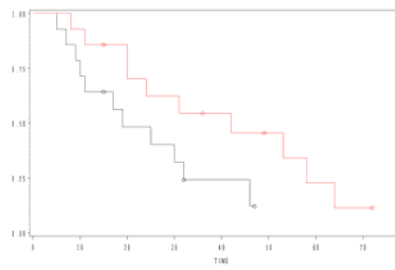
time	censor	TRT
179	1	1
378	0	0
256	1	0
355	1	1
262	1	1
319	1	0
256	0	0
256	1	1
255	1	1

Time: 投与後の生存時間  
 Censor: 1=打ち切りなし 0=打ち切り  
 TRT: 1=抗がん剤 0=投与なし  
 Sex1=男性 0=女性

抗がん剤を用いた場合と用いない場合・・・どっちが長生きするだろう？

生存時間解析(7)

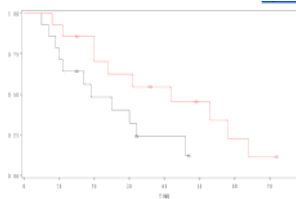
グラフの形だけは覚えておきましょう！



有名なKaplan-Meier 推定量曲線と呼びます！○印は打ち切りですが、打ち切りにより生存率が低下していないところに注目願います

生存時間解析(8)

複数群の生存時間に差があるかは・・・log-rank検定！



この場合、赤い曲線の群と黒い曲線の群には、生存率に差があるのだろうか？

帰無仮説は「0以上の全てのTIMEにおいて生存率は等しい」であり、対立仮説は「0以上のあるTIMEにおいて生存率は等しくない」である

最終的に生存率が10%程度であっても、赤い曲線の方が長生きしているように見える。だけど、治療法や薬剤による決定的な差なのか偶然なのかは分からない

生存時間解析(9)

黒い曲線群

赤い曲線群

層1: TRT = 0					層2: TRT = 1				
TIME	生存率	死亡率	生存率の標準誤差	死亡数	TIME	生存率	死亡率	生存率の標準誤差	死亡数
0.0000	1.0000	0	0	0	0.0000	1.0000	0	0	0
5.0000	0.9375	0.0625	0.0625	1	5.0000	0.9375	0.0625	0.0625	1
10.0000	0.8750	0.1250	0.0625	2	10.0000	0.8750	0.1250	0.0625	2
15.0000	0.8125	0.1875	0.0625	3	15.0000	0.8125	0.1875	0.0625	3
20.0000	0.7500	0.2500	0.0625	4	20.0000	0.7500	0.2500	0.0625	4
25.0000	0.6875	0.3125	0.0625	5	25.0000	0.6875	0.3125	0.0625	5
30.0000	0.6250	0.3750	0.0625	6	30.0000	0.6250	0.3750	0.0625	6
35.0000	0.5625	0.4375	0.0625	7	35.0000	0.5625	0.4375	0.0625	7
40.0000	0.5000	0.5000	0.0625	8	40.0000	0.5000	0.5000	0.0625	8
45.0000	0.4375	0.5625	0.0625	9	45.0000	0.4375	0.5625	0.0625	9
50.0000	0.3750	0.6250	0.0625	10	50.0000	0.3750	0.6250	0.0625	10
55.0000	0.3125	0.6875	0.0625	11	55.0000	0.3125	0.6875	0.0625	11
60.0000	0.2500	0.7500	0.0625	12	60.0000	0.2500	0.7500	0.0625	12
65.0000	0.1875	0.8125	0.0625	13	65.0000	0.1875	0.8125	0.0625	13
70.0000	0.1250	0.8750	0.0625	14	70.0000	0.1250	0.8750	0.0625	14

共通するTIME、もしくは近寄っているいくつかのTIMEにおける生存率を確認してみますと、差がありそうに見えなくもありません。実際に検定してみますと、

層に対しての同等性の検定

検定	カイ 2乗	自由度	Pr > Chi-Square
ログランク	3.4221	1	0.0643

p>0.05と有意ではありませんでした

比例ハザードモデル(1)

生命あるものは必ずいつかは死に、形あるものもいつか必ず壊れる運命である。当然であるが、若返ることも、新しくなることも不可能である

不健康な生活を続けていれば本来よりも寿命は短くなりやすく、粗末に扱えば壊れやすくなる。では、何が「寿命」に対して影響を与えているのか？

例:手術の有無、投薬の有無、喫煙の有無、飲酒の有無、稼ぎの良さ・・・？

これらの有無で層別して比較するのであれば、先ほどのlog-rank検定でよいのだけれども・・・？

比例ハザードモデル(2)

log-rank検定は1つの要因の差を見ることしかできない。寿命を決定する要因は必ずしも1つだけではない！むしろ1つであることはほとんどないだろう



ではロジスティック回帰は・・・？ 複数要因の影響を見ることは出来るが、「観察期間の終了までにある現象が発生するか否か」のみに注目した手法であり、それが発生するまでの時間は考慮されていない

log-rank検定では、ある一つの要因について差の有無がわかるだけであり、複数の要因について確認することは出来ない

多変量を見つ、さらに発生までの時間を考慮したスグレモノとして、Coxの比例ハザードモデルが存在する！

### 比例ハザードモデル(3)

そもそもハザード(Hazard)って何だろう？



ゴルフのハザードは「障害物」に近く、日本語では「危険」に近い意味で用いられている？

で、少々強引ですが「生存している=危険に遭遇する可能性がある」ということで、死亡してしまう率、ここでは「瞬間死亡率」と考える。このハザードを用いたモデルが「Coxの比例ハザードモデル」である！

瞬間死亡率とは、ある時間まで生きていた人が、次の瞬間に死亡する確率のことである

### 比例ハザードモデル(4)

ハザード(Hazard)関数は $h(t)$ で示される！

$$h(t) = h_0(t) \exp(a_1 \times \text{説明変数1} + \dots + a_m \times \text{説明変数}m)$$

\*exp(カッコ内) とは $e^{\text{カッコ内}}$ という意味です！

$$e = 2.718281828459014 \dots \text{(ここまでしか覚えてません)}$$

PATID	追跡期間	説明変数1	...	説明変数m	追跡終了時点の状況
患者1	t1	8.3	...	1(あり)	1(死亡)
患者k	tk	...	...	...	0(生存)
患者n	tn	-100	...	0(なし)	0(生存)

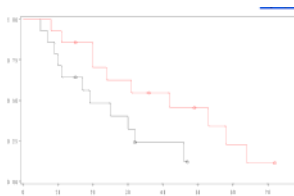
たとえば、患者1の時間tにおけるハザード関数は

$$h_1(t) = h_0(t) \exp(a_1 \times 8.3 + \dots + a_m \times 1)$$

説明変数が全て0の場合は、 $h(t) = h_0(t) \exp(a_1 \times 0 + \dots + a_m \times 0)$   
 $h(t) = h_0(t) \exp(0) = h_0(t) \times 1 = h_0(t)$ となります。これを**ベースラインハザード**と呼びます！  
 (復習:どんな数値でも0乗=1 です)

### 比例ハザードモデル(5)

ハザード(Hazard)比とはハザードの比、すなわち「ある時点まで生き延びた者が、その時点において死亡する確率(瞬間死亡率)」の比率のことである



比例ハザードモデルは、その比率が一定(=観察期間内のハザード比が一定)であるという仮説の下で解析が行われなければならない！

たとえば、左のグラフにおいては、常に両群の生存確率が平行に近い。ところが、もしもどこかで**交差**してしまうようでは、比例ハザード性の仮定が崩れてしまう

### 比例ハザードモデル(6)

では、どのように用いるのか？たとえば下記のデータ

PATID	生存期間(日)	打ち切り	年齢	性別	投与薬剤
患者1	542	あり	63	男	プラセボ
患者2	3087	なし	43	女	抗がん剤
...	...	...	...	...	...
患者n	936	あり	70	男	抗がん剤

あるがんの延命効果についてプラセボ群と抗がん剤群の投与群の比較試験を行った。生存時間は手術日～死亡までの日数である。薬剤の他、年齢や性別などの生存時間への影響を調べてみよう！

筆算では辛いので、ソフトウェアにお願いしてみましょう！  
 ここではSASを用います

### 比例ハザードモデル(7)

結果の見方について学びましょう！

PHREG プロシジャ  
最尤推定量の分析

変数	自由度	パラメータ推定	標準誤差	カイ2乗	P > ChiSq
AGE	1	0.00101	0.00719	0.0196	0.8885
SEX	1	-0.18328	0.17075	1.1922	0.2801
TREAT	1	-0.75733	0.17975	17.9514	<.0001

最尤推定量の分析

変数	ハザード比	95%ハザード比信頼区間
AGE	1.001	0.987 1.015
SEX	0.839	0.598 1.169
TREAT	0.489	0.330 0.699

p値はTREAT(抗がん剤orプラセボ)以外では有意ではありません。年齢や性別はそれほど関係ないってことですね。実際に、ハザード比が年齢はほぼ1倍に対し、TREATは半分以下です。**95%信頼区間も1をまたいでおりますし・・・？**

### 比例ハザードモデル(8)

OBS \_TIES\_ \_TYPE\_ \_STATUS\_ \_NAME\_ AGE SEX TREAT \_LNLIKE\_  
 1 BRESLOW PARMS 0 Converged TIME 0.001008930 -0.18328 -0.75733 732.291

というわけで、この場合のモデルは

$$h(t) = h_0(t) \exp(0.00101 \times \text{年齢} - 0.18328 \times \text{性別} - 0.75733 \times \text{抗がん剤})$$

となります(男=1、女=0、プラセボ=0、抗がん剤=1としてデータセットを作成しております)

符号から判断すると、年齢がひとつ上がる**とほんの少しだけ**死亡リスクが上がり、男性の方が**若干**リスクが下がり、抗がん剤を用いる**とリスクが大きく下がる**ということです！

結論:抗がん剤しか効いていないと見るのが妥当！

ちょっとだけ宣伝

弊社は「医師のYahoo」と呼ばれている(らしい?)エムスリー株式会社(m3.com)のグループ企業です! 医師の皆様の75%が登録されておりますので、臨床研究へのご協力やリクルートなどが、非常に迅速かつスムーズに行えます

詳細は <http://www.mebix.co.jp/> まで!

臨床研究のご相談などは右記まで! [ohashi@mebix.co.jp](mailto:ohashi@mebix.co.jp)

東京医科歯科大学医学部保健衛生学科大学院はこちら!  
[wohashi@bioinfo.tmd.ac.jp](mailto:wohashi@bioinfo.tmd.ac.jp)