

医学統計セミナー ベーシックコース
生存時間・臨床検査データの解析

下川 敏雄

和歌山県立医科大学 臨床研究センター

2016年度 医学統計セミナー

■ ベーシック・コース

- 基礎統計学 (6月15日・住金棟5F 大研修室)
- 量的データの解析 (7月27日・住金棟5F 大研修室)
- 質的データの解析 (8月24日・住金棟5F 大研修室)
- 共変量調整を伴う解析 (11月2日・病院棟4F 臨床講堂1)
- 生存時間・臨床検査データの解析(11月16日・住金棟5F 大研修室)

■ アドバンス・コース

- 多群・経時データの解析と多重比較
(11月30日・病院棟4F 臨床講堂1)
- 臨床試験における症例数設定とガイドライン
(12月28日・住金棟5F 大研修室)
- アンケート調査データの解析 (2月1日・病院棟4F 臨床講堂1)
- 統計的因果推論と傾向スコア (2月22日・住金棟5F 大研修室)
- メタアナリシス (3月22日・病院棟4F 臨床講堂1)

本講義の概要

• 生存時間解析

- 生存時間曲線の推定：Kaplan-Meier曲線
- 生存時間曲線の比較
- 最近の話題：競合リスク

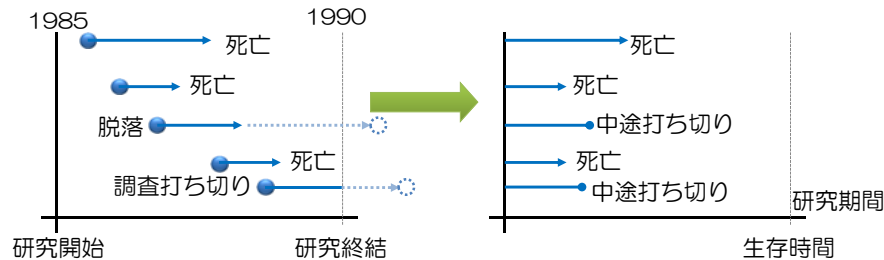
• 臨床検査データの解析

- 質的アウトカムに対する評価
- 量的・順序アウトカムに対する評価

生存時間解析

生存時間データの特徴

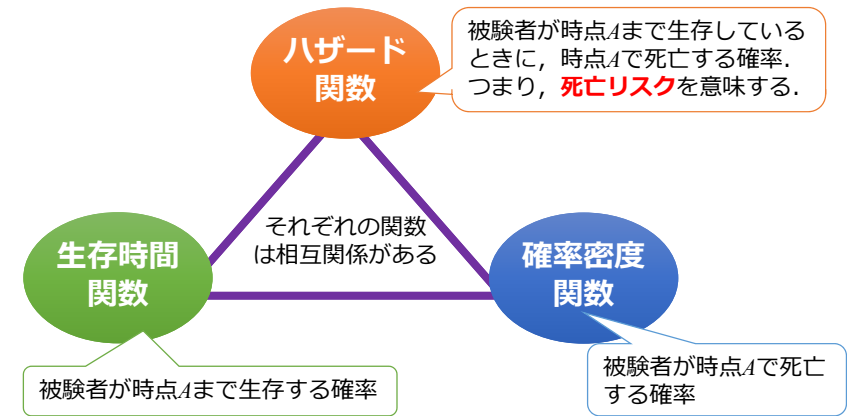
生存時間研究では、殆どの場合、中途打ち切り観測値や欠測値といった不完全データを含んでいる。



生存時間データは、個体の死亡(あるいは故障)までの時間を観測して得られる。観測される生存時間は生存時間(確率変数) T の実現値 t であるか、あるいは中途打ち切りが生じたときには、その中途打ち切り(打ち切りともいう)(censoring)時点の値である。

本報告でとり扱う右側中途打ち切りとは、個体の登録時点はわかるものの、追跡期間中に死亡(event)が起こらなかつたことを表す。

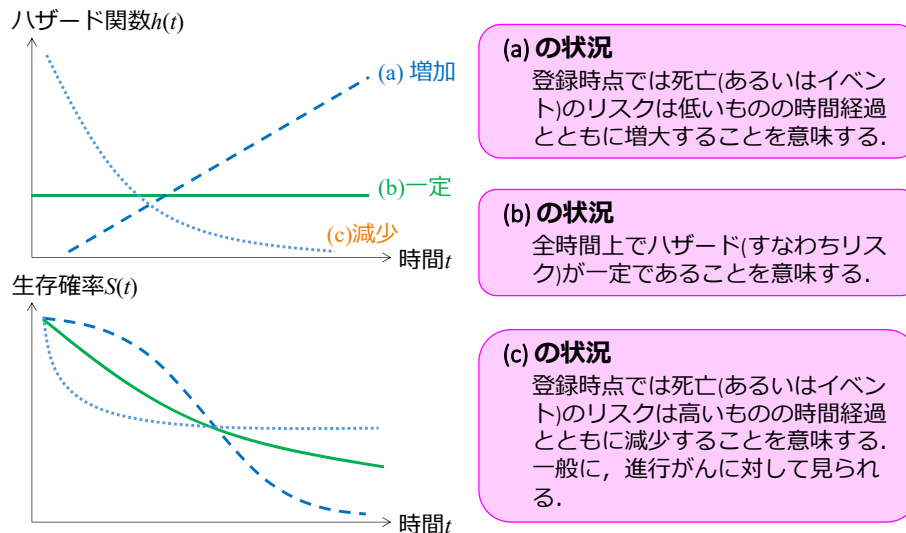
生存時間解析において重要な関数



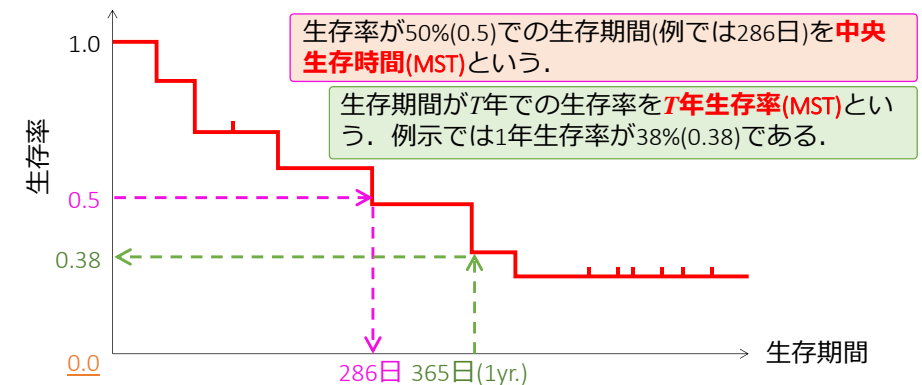
確率密度関数, 確率密度関数, 生存時間関数は相互関係があり, どれか1個が決まれば, すべてが決定できる。

ログランク検定ではハザードを評価しているのに、生存時間曲線を描写してグラフを描写している。これは、3種類の関数は相互関係があるため。

ハザード関数と生存時間関数の関係



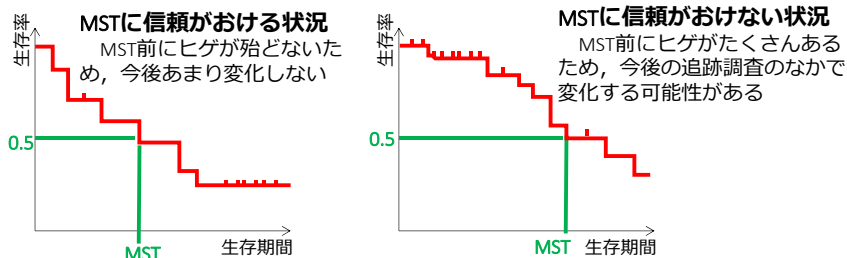
生存時間曲線の推定: Kaplan-Meier法



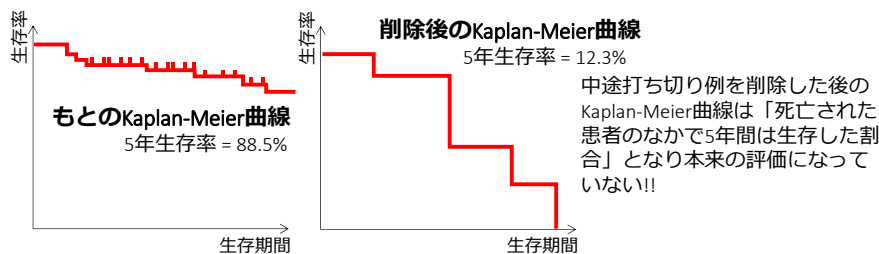
ここで注意点は、365日のところで、矩形の上側を見るのか下側を見るのかという点である。Kaplan-Meier曲線の矩形は上側ではなく下側での生存率を見なければならぬ。

Kaplan-Meier曲線の解釈上の留意点

■ 中途打ち切りを表すヒゲには意味がある



■ 中途打ち切り例を削除してはならない

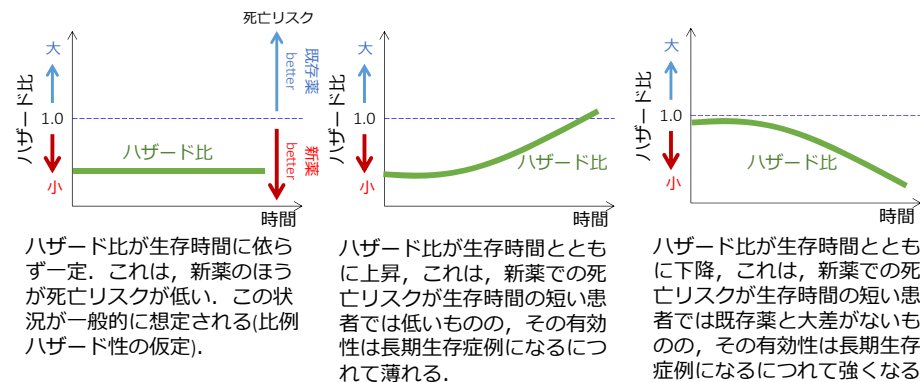


生存曲線の比較において重要な指標：ハザード比

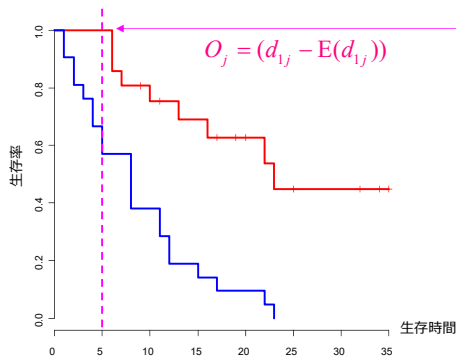
ハザード(危険度) $\lambda(t)$ とは「時間 t まで患者が生存しているという条件のもとで、時間 t で死亡する確率」を表す。つまり、時間 t での死亡リスクである。

ハザード比(相対危険度)とはハザードの比である。例えば、既存薬のハザードに対する新薬のハザードでのハザード比は

$$\text{ハザード比 } HR(t) = \frac{\text{新薬でのハザード } \lambda_1(t)}{\text{既存薬でのハザード } \lambda_0(t)}$$



ログランク検定とは何をしているか？



ある時点 t_j における死亡・生存に対するクロス集計表

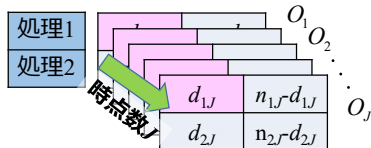
| | 死亡 | 生存 | 合計 |
|-----|----------|-------------------|----------|
| 処理1 | d_{1j} | $n_{1j} - d_{1j}$ | n_{1j} |
| 処理2 | d_{2j} | $n_{2j} - d_{2j}$ | n_{2j} |
| 合計 | d_j | $n_j - d_j$ | n_j |

もし、処理1と処理2で死亡数が同じであれば(つまり処理効果がなければ)、 \square の期待度数は、カイ2乗検定と同じ理屈から

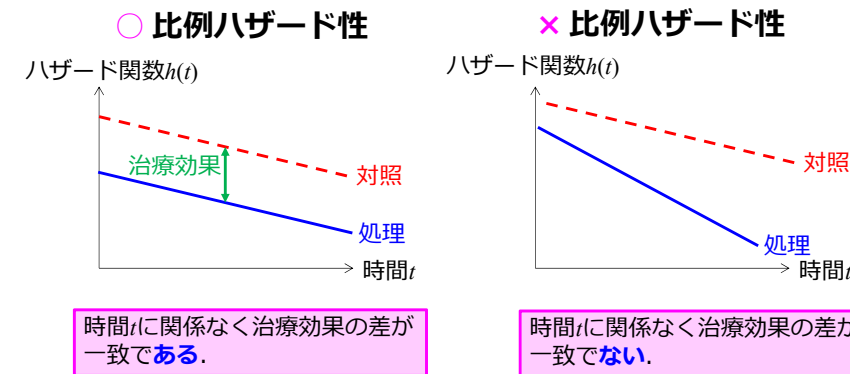
$$E(d_{1j}) = \frac{d_j n_{1j}}{n_j}$$

である。期待度数からの乖離

上記の2x2クロス集計表の \square 部分をすべての時点において計算し、その総和をとる $O_1 + O_2 + \dots + O_j$ ことで構成されたのが、ログランク検定である。



ログランク検定における仮定：比例ハザード性



比例ハザード性の仮定は、ログランク検定だけでなく比例ハザードモデルにおいてもおかれている。

もうひとつの検定方法：一般化Wilcoxon検定



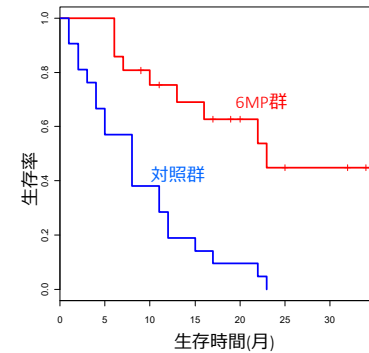
時間が小さい場合には、 $\hat{S}(t_j)$ が1に近い(重みが大きい)ことから、一般化Wilcoxon検定では、早期の2群間の差に敏感である。

2標本の比較の例示(1)

小児急性白血病データ
(Freireich *et al.*, 1963)

| | | | | | | | | |
|------------|-----|----|----|-----|-----|-----|----|-----|
| 6MP 投与群 | 6 | 6 | 6 | 6+ | 7 | 9+ | 10 | 10+ |
| | 11+ | 13 | 16 | 17+ | 19+ | 20+ | 22 | 23 |
| 対照群 | 1 | 1 | 2 | 2 | 3 | 4 | 4 | 5 |
| | 5 | 8 | 8 | 8 | 8 | 11 | 11 | 12 |
| | 12 | 15 | 17 | 22 | 23 | | | |

+は中途打ち切り, 単位は月



諸種の検定におけるp値

ログランク検定 $p < 0.0001$
 Gehanの一般化Wilcoxon検定 $p = 0.0002$
 Peto-Petoの一般化Wilcoxon検定 $p = 0.0002$

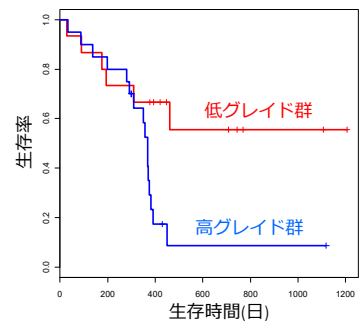
いずれの検定法でも6MP群と対照群のあいだで有意であることがわかる。

2標本の比較の例示(2)

卵巣がん患者のグレード毎での病気進行期間(Fleming *et al.*, 1980)

| | | | | | | | | |
|--------|------|------|------|-------|------|-------|-------|------|
| 低グレード群 | 28 | 89 | 175 | 195 | 309 | 377+ | 393+ | 421+ |
| | 447+ | 462 | 709+ | 744+ | 770+ | 1106+ | 1206+ | |
| 高グレード群 | 34 | 88 | 137 | 199 | 280 | 291 | 299+ | 300+ |
| | 309 | 351 | 358 | 369 | 369 | 370 | 375 | 382 |
| | 392 | 429+ | 451 | 1119+ | | | | |

+は中途打ち切り, 単位は日



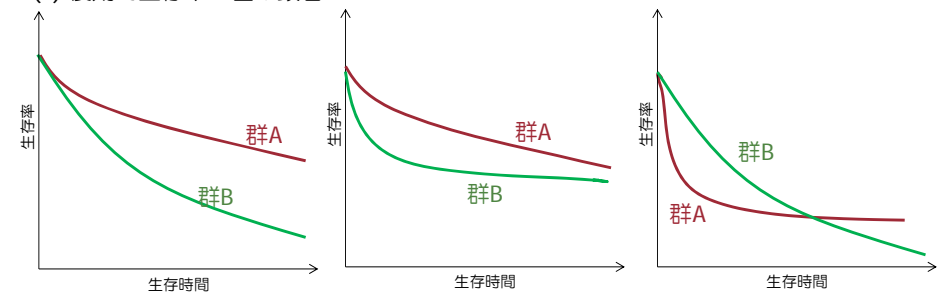
諸種の検定におけるp値

ログランク検定 $p = 0.0183$
 Gehanの一般化Wilcoxon検定 $p = 0.0995$
 Peto-Petoの一般化Wilcoxon検定 $p = 0.1342$

比例性は満たされていないと指摘されているデータではあったものの、ログランク検定のみで有意差が認められた。初期の生存時間で差が認められなかったことが寄与していると考えられる。

手法による違い(大橋・浜田,1995)

(a) 後期で生存率の差が顕著 (b) 前期で生存率の差が顕著 (c) 生存曲線が交差

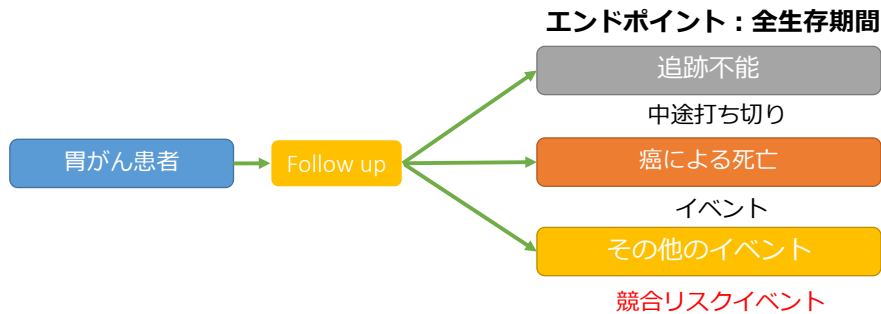


(a)の状況では、ログランク検定のほうが一般化Wilcoxon検定よりも検出力が強い。
 (b)の状況では、一般化Wilcoxon検定のほうがログランク検定よりも検出力が高い。
 (c)の状況では、いずれの方法も検出力が低い。

検出力とは、治療効果があるときに「有意である」と正しく判断できる確率を意味する。

最近の調査研究で用いられる方法：競合リスク

競合リスクの例示：高齢者の胃がん患者に対する抗癌剤の評価



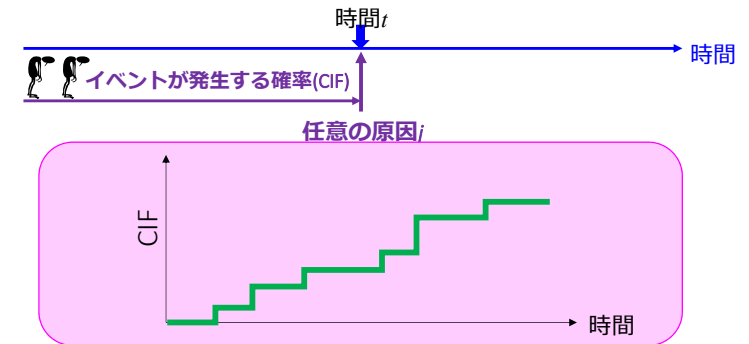
高齢者のがん患者では、がんによる死亡よりも前に、その他の要因(例えば、循環器疾患)による死亡が起きる可能性が高い。このとき、全生存期間は抗癌剤の有効性を十分に反映しない結果となり得る。

臨床検査データの解析

競合リスクでは生存率ではなく発生率を評価する

■ 累積発生関数(CIF; Cumulative Incident Function)

原因 j によるイベントが時間 t までに発現する確率(Kalbfleisch & Prentice, 1980)



通常の生存時間解析では、

生存率：任意の生存期間までに、生存(イベントがない)している割合を評価するが、競合リスクの場合にはイベントにラベリングが行われているので、

累積発現率：任意の生存期間までに、任意のイベントが発現する割合を評価する。

カテゴリカルデータ(定性テスト)の評価

| | | 疾病の有無 | | 特異度 |
|------|----|---------|---------|-----|
| | | 有 | 無 | |
| 検査結果 | 陽性 | 真陽性(TP) | 偽陽性(FP) | ↓ |
| | 陰性 | 偽陰性(FN) | 真陰性(TN) | |

疾患の有無には有病率が関係する

- 真陽性(TP)：疾患有の被験者を陽性と正しく診断すること。
- 偽陰性(FN)：疾患有の被験者を陰性と誤って診断すること。
- 真陰性(TN)：疾患無の被験者を陰性と正しく診断すること。
- 偽陽性(FP)：疾患無の被験者を陽性と誤って診断すること。

感度(Sensitivity)

感度 = $TP / (TP + FN)$ ：疾患有の被験者を陽性と正しく判断する確率

特異度(Specificity)

特異度 = $TN / (TN + FP)$ ：疾患無の被験者を陰性と正しく判断する確率

感度・特異度の留意点

感度(Sensitivity)

感度 = TP/(TP+FN) : 疾患有の被験者を陽性と正しく判断する確率

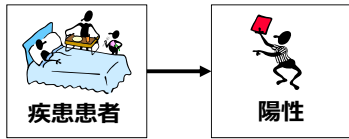
特異度(Specificity)

特異度 = TN/(TN+FP) : 疾患無の被験者を陰性と正しく判断する確率

利点 : 当該疾患の有病率(prevalence)に影響されず診断性能を評価できる。

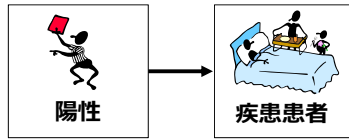
注意点 : 「感度 ≠ 疾患の確率」であることに注意する。

感度



感度は、結果がわかっている状況のなかで評価している。

臨床的有用性



陽性と診断されたときに、本当に疾患だった確率が重要になる。

陽性的中率(陽性予測値)・陰性的中率(陰性予測値)

| | | 疾病の有無 | | |
|------|----|---------|---------|-------|
| | | 有 | 無 | |
| 検査結果 | 陽性 | 真陽性(TP) | 偽陽性(FP) | 陽性的中率 |
| | 陰性 | 偽陰性(FN) | 真陰性(TN) | 陰性的中率 |

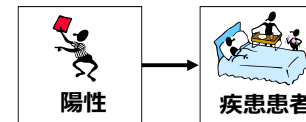
疾患の有無には有病率が関係する

陽性的中率(positive predictive value)

陽性的中率 = TP/(TP+FP) : 陽性の被験者のなかで疾患有の確率

陰性的中率(Specificity)

陰性的中率 = TN/(TN+FN) : 陰性の被験者のなかで疾患無の確率



利点 : 臨床的有用性は陽性的中率で評価できる。

注意点 : これらの指標は、有病率の影響を受ける。

感度・特異度と陽性的中率・陰性的中率の関係

感度・特異度と陽性的中率・陰性的中率の関係は、Bayesの定理を用いて計算できる。

$$\text{陽性的中率} = \frac{\text{有病率} \times \text{感度}}{(\text{有病率} \times \text{感度}) + (1 - \text{有病率}) \times (1 - \text{特異度})}$$

この数式からも陽性的中率では有病率が問題になることがわかる。

余談



Thomas Bayes(1702-1761)

Bayesの定理とは、Thomas Bayesによって発見された定理である。このBayesの定理を用いる統計学のことをBayes統計学という。

その利点は、データ以外の知識・経験などの様々な情報を事前情報として、未治療の推測のために活用できる。また、新たなデータを得たときには、これを事前情報(事前分布)として取り込むことができる。

Bayesの定理そのものは、300年近くまえに発見されたものだが、近年Bayes統計学が脚光を浴び、医学・薬学分野にも広く応用されている。

診断精度の評価尺度：尤度比

| | | 疾病の有無 | |
|------|----|-----------|------------|
| | | 有 | 無 |
| 検査結果 | 陽性 | 感度(TPR) | 1-特異度(FPR) |
| | 陰性 | 1-感度(FNR) | 特異度(TNR) |

陽性尤度比(likelihood ratio of a positive result)

$$\text{陽性尤度比} = \text{TPR}/\text{FPR} = \text{感度}/(1 - \text{特異度})$$

「疾患有が疾患無よりも何倍陽性になりやすいか」を表す指標

疾患有を陽性と診断した場合と疾患無を陽性と診断した場合の比で表されており、大きいほど確定診断に優れるといえる(一般に尤度比といった場合には陽性尤度比を表す。)

陰性尤度比(likelihood ratio of a negative result)

$$\text{陰性尤度比} = \text{FNR}/\text{TNR} = (1 - \text{感度})/\text{特異度}$$

「疾患有が疾患無よりも何倍陰性になりやすいか」を表す指標

疾患有を陰性と診断した場合と疾患無を陰性と診断した場合の比で表される指標である(あまり用いられない)。

例示：疑似データ(新谷, 2016)

| | | 疾病の有無 | | 合計 |
|------|----|------------|-------------|-------------|
| | | 有 | 無 | |
| 検査結果 | 陽性 | 24 (TP) | 19 (FP) | 43 (TP+FP) |
| | 陰性 | 16 (FN) | 941 (FP) | 957 (FN+FP) |
| 合計 | | 40 (TP+FN) | 960 (FP+FP) | 1,000 |

感度(Sensitivity)

感度(TPR) = $TP / (TP + FN) = 24 / 40 = 0.60$

特異度(Specificity)

特異度(TNR) = $TN / (TN + FP) = 941 / 960 = 0.98$

陽性的中率(positive predictive value)

陽性的中率 = $TP / (TP + FP) = 24 / 43 = 0.56$

陰性的中率(Specificity)

陰性的中率 = $TN / (TN + FN) = 941 / 957 = 0.98$

例示

| | | 疾病の有無 | | 合計 |
|------|----|------------|-------------|-------------|
| | | 有 | 無 | |
| 検査結果 | 陽性 | 24 (TP) | 19 (FP) | 43 (TP+FP) |
| | 陰性 | 16 (FN) | 941 (FP) | 957 (FN+FP) |
| 合計 | | 40 (TP+FN) | 960 (FP+FP) | 1,000 |

TPR = 0.60, TNR = 0.98

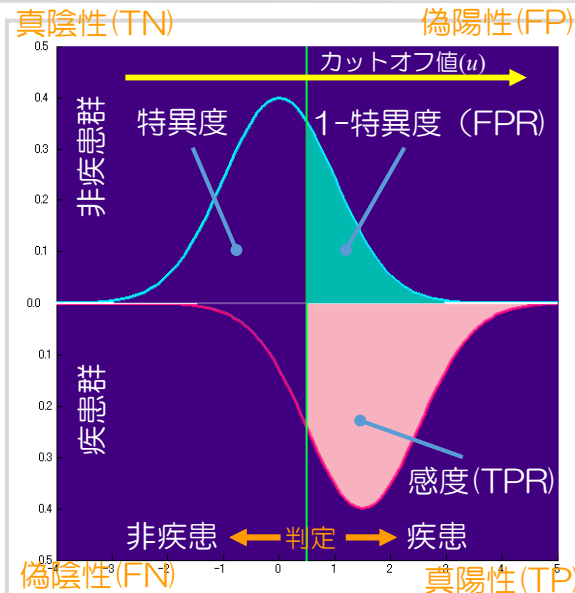
陽性尤度比(likelihood ratio of a positive result)

陽性尤度比 = $TPR / FPR = 0.60 / (1 - 0.98) = 30.32$

陰性尤度比(likelihood ratio of a negative result)

陰性尤度比 = $FNR / TNR = (1 - 0.60) / 0.98 = 0.41$

量的データ・順序データの評価：ROC曲線



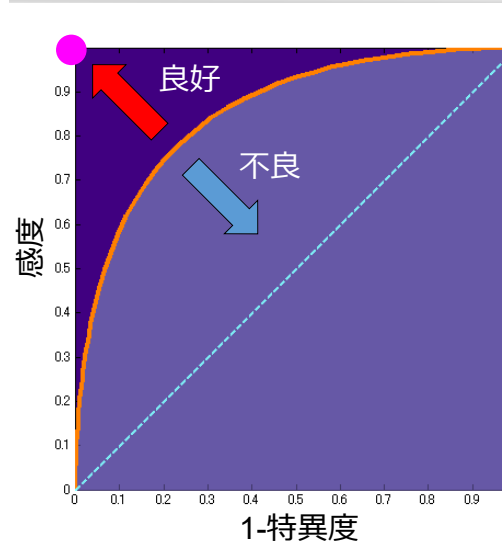
真陽性
疾患群において、カットオフ値(u)以上となる。

偽陽性
疾患群において、カットオフ値(u)未満となる。

真陰性
非疾患群において、カットオフ値(u)以上となる。

偽陰性
非疾患群において、カットオフ値(u)未満となる。

量的データ・順序データの評価：ROC曲線



ROC曲線

カットオフ値uを小さいほうから大きいほうに動かしたときの、(1-特異度, 感度)をプロットしたものがROC曲線である。

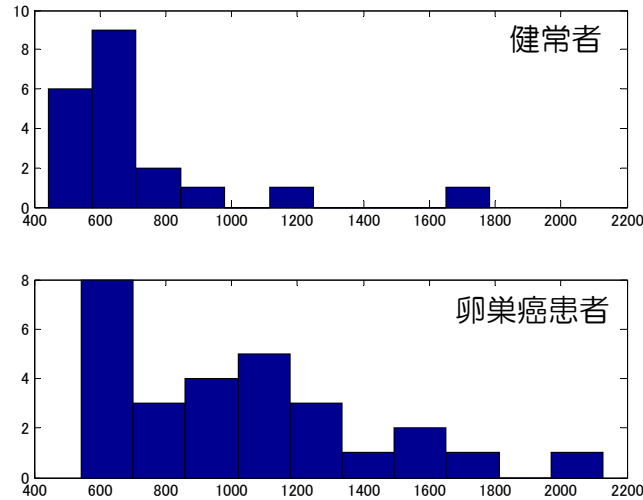
ROC曲線は、45度の直線に対して扇型の形状を示す。座標(0,1)[●で表された場所]に近くなるほど良好な診断であるといえる。

曲線化面積AUC

ROC曲線の曲線以下の面積である(■の面積)。複数の検査を比較したり、診断能を数値化するのに用いられる。

事例：卵巣癌データ(Pepe,2003)

卵巣癌の有無を識別するために、卵巣癌患者30名および健常者23名の遺伝子データが観測されている(Pepe, 2003). 本解析の目標は、任意の遺伝子が卵巣癌の発現に寄与しているか否かを評価することにある。



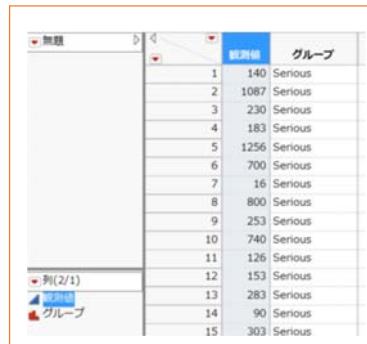
事例に対するROC曲線

■頭部外傷症データ(Zhou, et al.,2002)

頭部外傷症の重篤度を識別するために、CK-BB(クレアチン・キナーゼBB)が有効か否かを判定している。ここに、重篤度は、重度および非重度の2値とする。Zhou et al.(2002)は、CK-BBが成人での重篤度の判定だけでなく、未成人においても利用可能であるか否かを検討している。

| 重症群 | | | | | 非重症群 | | |
|------|------|-----|------|-----|------|-----|----|
| 140 | 740 | 543 | 490 | 523 | 136 | 60 | 46 |
| 1087 | 126 | 913 | 156 | 76 | 286 | 17 | |
| 230 | 153 | 230 | 356 | 303 | 281 | 27 | |
| 183 | 283 | 463 | 350 | 353 | 23 | 126 | |
| 1256 | 90 | 60 | 323 | 206 | 200 | 100 | |
| 700 | 303 | 509 | 1560 | | 146 | 253 | |
| 16 | 193 | 576 | 120 | | 220 | 70 | |
| 800 | 76 | 671 | 216 | | 96 | 40 | |
| 253 | 1370 | 80 | 443 | | 100 | 6 | |

JMPによる実行



JMPの手順

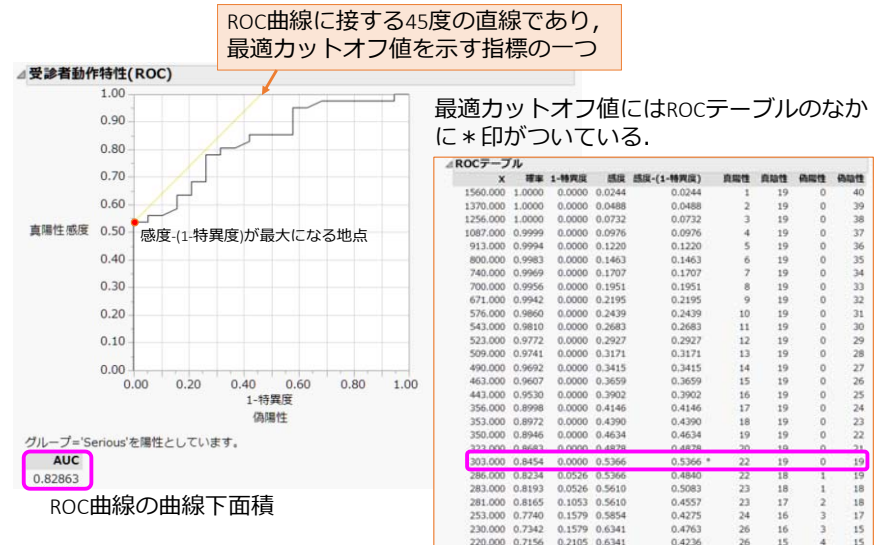
STEP.1 「分析」 → 「2変数の関係」 ボタンを押す。

STEP.2 「Y, 目的変数」にグループ「X, 説明変数」に観測値を選択する。

(ロジスティック当てはめ) のウィンドウが表示される。

STEP.3 「▼」からROC曲線を選択する。

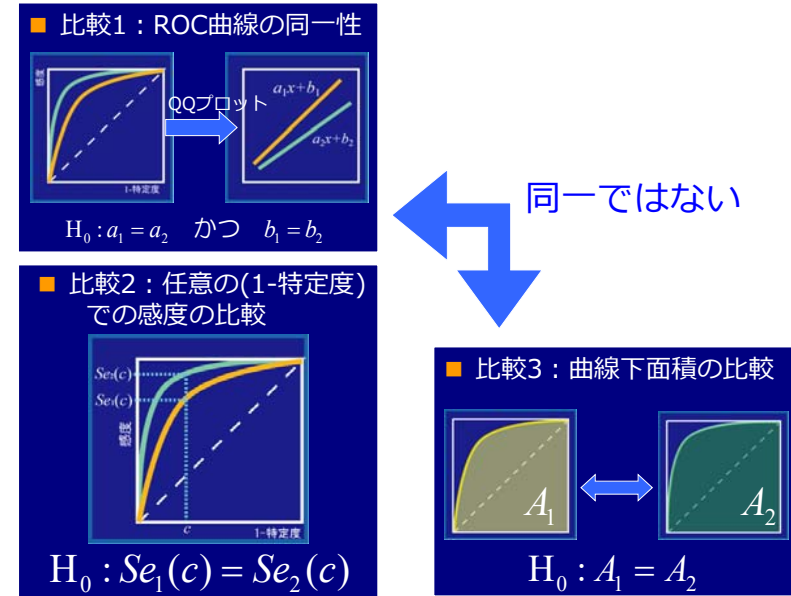
ROC曲線の結果



ROC曲線の比較：対応がある場合とない場合



ROC曲線の比較の諸型



JMPによる対応のあるデータでのROC曲線のAUCの比較

ROC曲線のAUCを比較するには、JMPの外部スクリプトを利用する必要があります。外部スクリプトは、臨床研究センターで管理しているため、要望がある場合は、連絡をお願いします。

JMPの手順

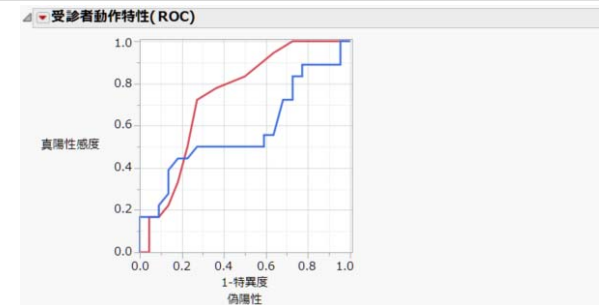
- STEP.1 データファイルを読み込む
- STEP.2 「ファイル」→「対応のあるROC曲線の比較」を選択し、「開く」の横の▼を押して実行を選ぶ。



疾患の有無を表す変数

検査値(2個以上)

結果



ROC曲線の曲線下面積及び95%CI

| 検査方法 | AUC | 標準誤差 | 下限95% | 上限95% |
|-------|--------|--------|--------|--------|
| 検査方法1 | 0.7361 | 0.0805 | 0.5533 | 0.8627 |
| 検査方法2 | 0.5783 | 0.0968 | 0.3864 | 0.7491 |

ROC曲線の曲線下面積の検定

| カイ2乗 | 自由度 | p値(Prob>ChiSq) |
|--------|-----|----------------|
| 3.7074 | 1 | 0.0542 |

ROC曲線の曲線下面積の差の95%CI

| 差 | 標準誤差 | 下限95% | 上限95% | カイ2乗 | p値(Prob>ChiSq) |
|---------------|--------|--------|---------|--------|----------------|
| 検査方法1 - 検査方法2 | 0.1578 | 0.0820 | -0.0028 | 3.7074 | 0.0542 |

ご清聴ありがとうございました

