

医学統計学演習：資料 14 簡易解答

芳賀昭弘 *

以下の演習問題のレポートを作成し **word** で haga@tokushima-u.ac.jp まで提出してください。

ファイル名は学籍番号_ローマ字.docx とすること。

例：2024000000_haga_akihiro.docx (**c** を入れない)

締め切り：2026 年 1 月 28 日深夜 23:59 まで

レポート作成にあたっては以下を留意すること。

- ・レポートは必ず自分でやること。他の人に見せるのも同罪で不正行為とみなします。また、生成 AI の丸写しはダメです。
- ・生成 AI を使うと、学んでいないライブラリや関数、表現を使って解答が書かれる場合があります。授業で使用していないライブラリや関数、表現を使う箇所があった場合、a) 授業ではどのようにしていたのか、b) なぜそれが必要なのか、を記述がなければ解答とみなしません。
- ・グラフの作成では、「図番号:タイトルとキャプション」を図の直下に記し、縦軸・横軸、レジェンドを挿入すること。また、単位がある量についてはその単位を明示すること。
- ・表の作成では、「表番号:タイトルとキャプション」を表の上に記すこと。単位がある量についてはその単位を明示すること。
- ・図表ともに略語は避けること。略語を使う場合はその前にフルスペルと略語の対応を示すこと（例：Standard Deviation (SD)）。
- ・レポートに定義されていない記号を使用しないこと（使用する場合はきちんと定義してから使用すること）。
- ・数字を必要以上に桁も書かないこと。有効数字があればその桁まで記載すること。

1 演習問題

1. E14-1.csv の診断結果値（無次元）データについて以下の問題に答えよ。

ここで、項目 A は健康、項目 B は病気の、各 50 人の AI による診断結果値、項目 C は健康、項目 D は病気の、各 50 人の医師による診断結果値、である。(1) 項目 A と項目 B についてヒストグラムを表示し、それぞれの平均、不偏標準偏差、中央値を有効数字 3 桁まで記載せよ。

(2) 同様に項目 C と項目 D についてヒストグラムを表示し、それぞれの平均、標準偏差、中央値を記載せよ。

(3) 項目 A と項目 B の平均値に差があるか検定せよ。ただし、有意水準を両側 1% する。また、それぞれの母集団は正規分布に従っているとする。

(4) 項目 A（健康）、項目 B（病気）の各 50 人の AI による診断結果値から、AI が病気と健康を分類で

* Electronic address: haga@tokushima-u.ac.jp

きるかを考えたい。そのために AI による診断結果値に基づいて ROC 曲線を作成し、AUC の値を算出せよ。

(5) 項目 C（健康），項目 D（病気）の各 50 人の医師による診断結果値から、医師が病気と健康を分類できるかを考えたい。そのために医師による診断結果値に基づいて ROC 曲線を作成し、AUC の値を算出せよ。

(6) 上記 (4),(5) について検討する。A と B のヒストグラムと (4) の ROC の曲線の形, C と D のヒストグラムと (5) の ROC の曲線の形について、その関係性を考察せよ。

(7) 偽陰性を極力減らしたい場合、上記の AI と医師とでどちらの診断結果値が良いと考えられるか。解答の根拠も述べること。

(8) (4) について、項目 A の平均値を閾値としたとき、項目 B の平均値を閾値としたときのそれぞれの 2×2 分割表を作成せよ。ここで、閾値以下は健康と判断するものとする。

(1)

A: 平均 2.86, 不偏標準偏差 0.924, 中央値 2.86

B: 平均 5.53, 不偏標準偏差 3.32, 中央値 5.59

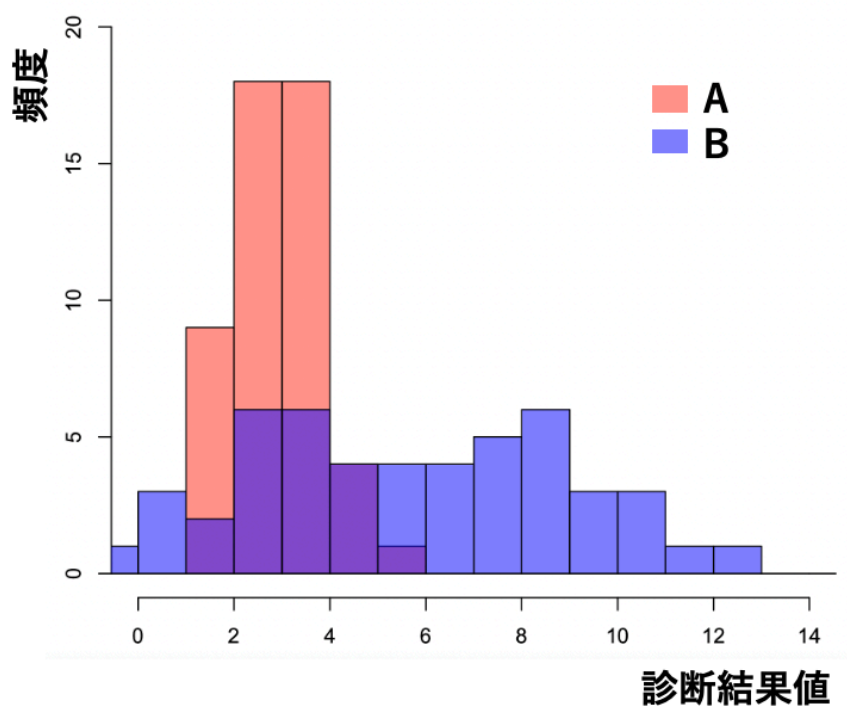


図1 項目 A（健康；赤）と項目 B（病気；青）の AI による診断結果値のヒストグラム。

(2)

C: 平均 3.84, 不偏標準偏差 3.13, 中央値 3.68

D: 平均 7.05, 不偏標準偏差 1.04, 中央値 7.16

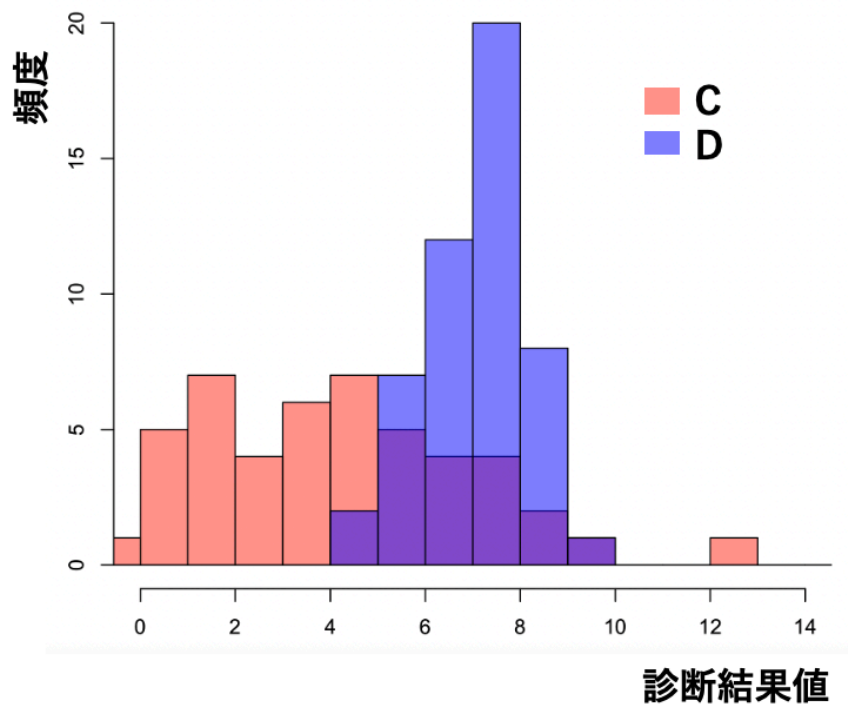


図2 項目 C（健康；赤）と項目 D（病気；青）の医師による診断結果値のヒストグラム。

(3)

項目 A と項目 B が対応しているか不明である。そのため対応がないものとして検定を進めるため、まずは項目 A と項目 B の分散に差があるかどうかを検定する。分散に差がある場合は Welch 検定、差がない場合は通常の対応のない t-検定を行う。

帰無仮説：項目 A と項目 B の分散に差がない。対立仮説：項目 A と項目 B の分散に差がある。

検定統計量：F-検定（母集団は正規分布に従う場合の 2 群の分散の差の検定であるため、F-検定を使用することができる。）

有意水準：両側 1%

統計量の実現値： $F = 0.0776$

p 値： $1.205e-15$

p 値が有意水準に対して十分小さいため、帰無仮説が棄却され対立仮説が採用される。つまり、項目 A と項目 B の分散に差がある。

分散に差があるため、項目 A と項目 B の平均値の差を検定するには Welch 検定を実施する。

帰無仮説：項目 A と項目 B の平均値に差がない。対立仮説：項目 A と項目 B の平均値に差がある。

検定統計量：t-検定（母集団は正規分布に従う場合の 2 群の平均値の差の検定であるため、t-検定を使用することができる。）

有意水準：両側 1%

統計量の実現値： $t = -5.48$

p 値： $1.009\text{e-}6$

p 値が有意水準に対して十分小さいため、帰無仮説が棄却され対立仮説が採用される。つまり、項目 A と項目 B の平均値に差がある。

(4)

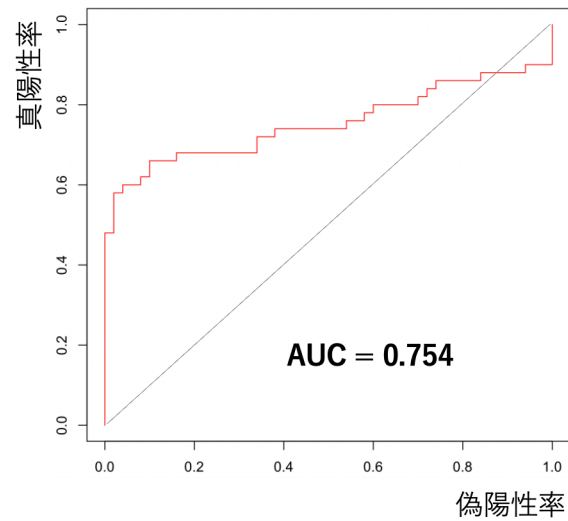


図 3 AI による健康（項目 A）と病気（項目 B）の診断結果値の Receiver Operating Characteristic (ROC) 曲線と Area Under the Curve (AUC)。

(5)

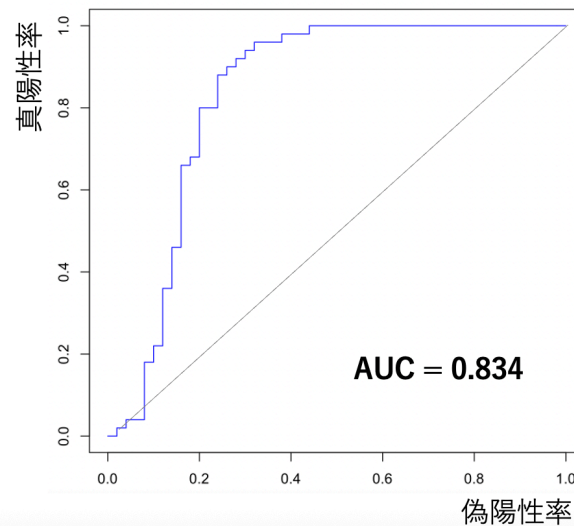


図 4 医師による健康（項目 C）と病気（項目 D）の診断結果値の Receiver Operating Characteristic (ROC) 曲線と Area Under the Curve (AUC)。

(6)

図 1 と図 2 のヒストグラムを見比べると、図 1 は健康な場合の分布の分散が小さく、病気の場合の分散が大きい、図 2 は健康な場合の分布の分散が大きく、病気の場合の分散が小さいという特徴を持っている。健康な場合の分布の分散が小さく、病気の場合の分散が大きい図 1 の場合、ROC 曲線の形としては左側に膨らんだ曲線形となる（図 3 のように偽陽性率が小さい段階で真陽性率が高くなる）。健康な場合の分布の分散が大きく、病気の場合の分散が小さい図 2 の場合、ROC 曲線の形としては右側に膨らんだ曲線形となる（図 4 のように偽陽性率が小さい段階で真陽性率はさほど高くない。一方、偽陽性率が 1 から小さくなっても真陽性率は 1 を保っている）。

(7)

偽陰性とは、本当は病気であるが誤って健康と診断することである。これは図 1 と図 2 のヒストグラムからその大きさがわかる。例えば閾値を 4 に設定した場合、図 1 のヒストグラムでは、病気 (B) では 4 より小さいところに分布する数が多い。この数は偽陰性数であるので、病気に対する診断結果値が広く分布するような図 1 の場合、偽陰性はある程度避けられないことを意味する。一方、図 2 のヒストグラムでは、病気 (D) は 4 より小さいところに分布していない。つまり偽陰性はこの場合 0 である。病気に対して診断結果値が狭い範囲で分布するような場合は、偽陰性を少なくできる可能性がある（D の分布がもっと左側によっていくとその限りではなくなる）。

従って、極力偽陰性を排除したいのであれば、今回は医師による診断結果値を採用すべきである。

(8)

表 1: A の平均値 2.86 を閾値に採用した場合の 2×2 分割表。

		実際		
		病気	健康	計
予測	陽性	37	25	62
	陰性	13	25	38
計		50	50	100

表 2: B の平均値 5.53 を閾値に採用した場合の 2×2 分割表。

		実際		
		病気	健康	計
予測	陽性	26	1	27
	陰性	24	49	73
計		50	50	100

2. E14.2.csv はがん患者 (id)、腫瘍マーカーの値 (maker)、生存期間 (survival_period, 単位:月)、死亡 (dead.status = 0: 打ち切り, dead.status = 1: 死亡) に関するデータが含まれている。このデータを使い、以下の問題に答えよ。

(1) 全データによる Kaplan-Meier 曲線を描け。

(2) 腫瘍マーカーの中央値で 2 群に分けて Kaplan-Meier 曲線を描き、腫瘍マーカーの値で生存期間に差があるかどうかを検定せよ。

(3) 生存期間 40 ヶ月の前後で患者を分けた場合、腫瘍マーカーの値で生存期間が 40 ヶ月以上となるか未満となるかを分けられるかどうかを判定する。腫瘍マーカーの値に対する生存期間の分類の ROC 曲線を描き、その AUC を求めよ。ただし、dead.status = 1 のみの症例を対象とせよ。

(1)

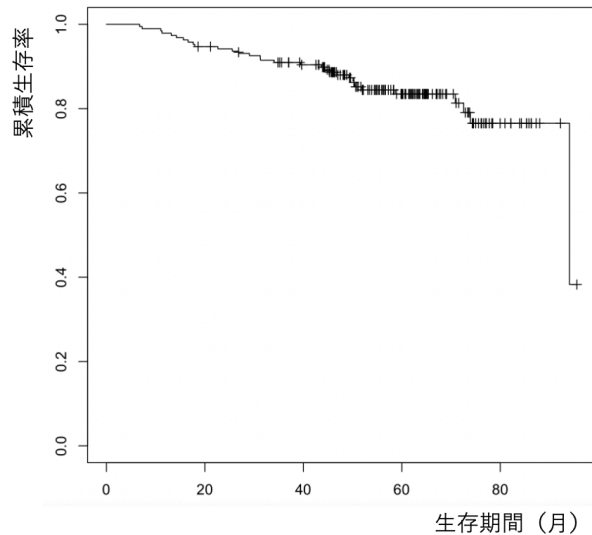


図5 E14_2.csv の全症例による Kaplan・マイヤー曲線。

(2) maker の中央値は 0.608。この前後でデータを分けると、maker の大きい方が生存期間が長いことがわかる (maker 大: 平均 63 ヶ月, maker 小: 平均 49 ヶ月)。

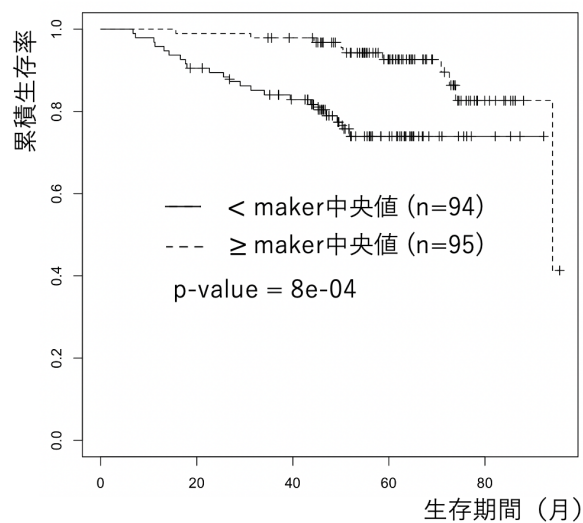


図6 腫瘍マーカーの中央値で分けた 2 群の Kaplan・マイヤー曲線; Solid line: 中央値未満, dashed line: 中央値以上。p-value は一般化 Wilcoxon 検定による結果である。

図に示したように meta の有無別の生存曲線に対する一般化 Wilcoxon 検定では $p \leq 0.0001$ であり、有意に差があると言える (腫瘍マーカーの中央値以上の方が生存率が高い)。

(3) dead_status = 1 の症例は 32 であり、そのうち生存期間が 40 ヶ月以上の症例数は 14 症例である。腫瘍マーカーの値に閾値を設けてその閾値以上の腫瘍マーカー値を持つものを陰性、未満を陽性と定義する。そうした時の ROC 曲線は図 7 のようになり、AUC は 0.794 となる。

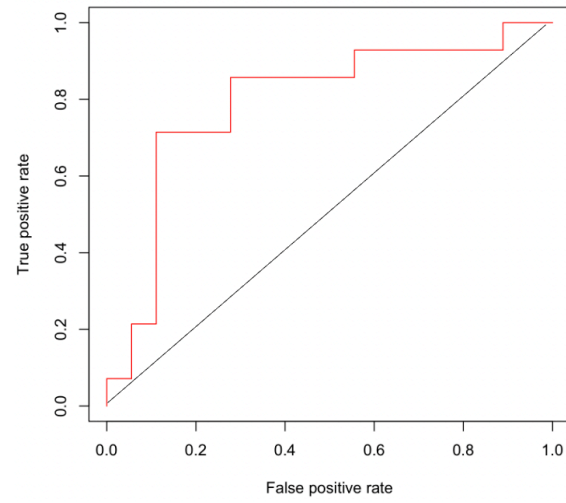


図7 生存期間が40ヶ月以上となるかどうかを、腫瘍マーカーの値で判定した時のROC曲線。