

ユークリッド距離と相関係数に基づく距離

文責: 中井尚一 (2022年 高桑研究室 M1)

最終更新日: 2022/06/09

<ユークリッド距離と相関係数に基づく距離の導出>

【ユークリッド距離】

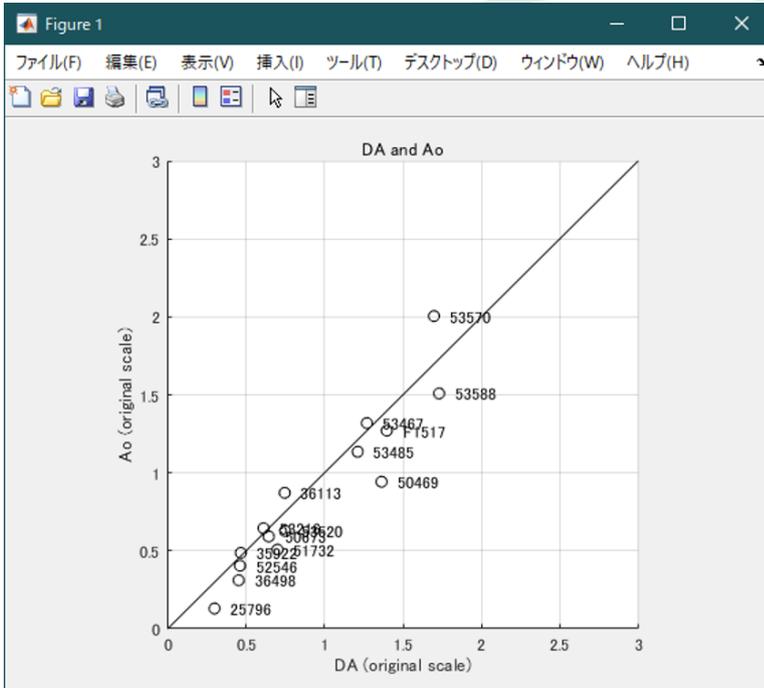


図1は横軸にDA（動脈管）の径、縦軸にAo（大動脈）の径をプロットした散布図である。また、 $y=x$ の直線との位置関係より、個体ごとのDAとAoの大小関係を捉えやすくしている。例えば、 $y=x$ の直線より上側にある53570の個体では、DAよりもAoの方が径が大きいといった具合である。

↑ 図1.DAの径とAoの径の散布図

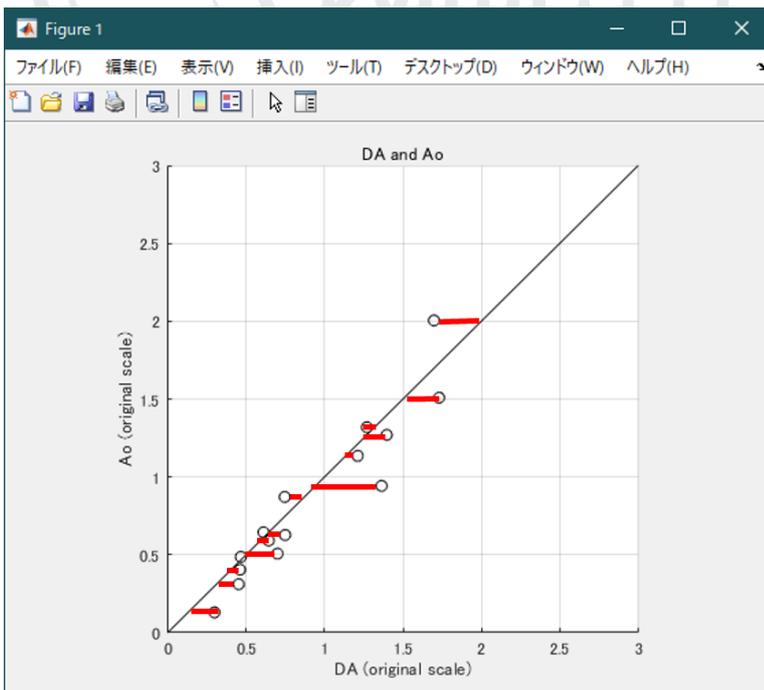
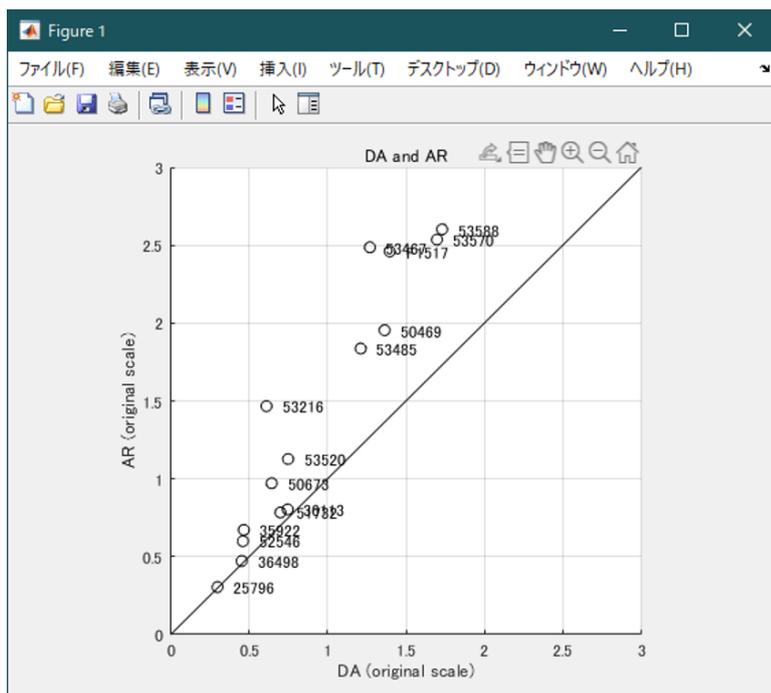


図1の各プロットから、 $y=x$ の直線に対して、図2で赤線で示したように足を伸ばす。このとき、それぞれの赤線は、各個体におけるDAの値とAoの値の差を示している。これを長さとして扱い、全個体分足し合わせたものがDAとAoにおけるユークリッド距離である。具体的には、DAとAoの差を2乗し、全て足し合わせて、平方根を取ることでユークリッド距離は導出される。

↑ 図2.各個体ごとのDAとAoの差

【相関係数に基づく距離】



↑図3.DAの径とARの径の散布図

図3は横軸にDAの径を、縦軸にAR（大動脈基部）の径をプロットした散布図である。ここで、DAの径とARの径の平均と偏差に着目すると、一致していないことがわかる。このままでは扱いにくいので、ノーマライズという操作を行う。具体的には、各値から平均を引き、標準偏差で割ることで、DAとARを共に平均0、標準偏差1で揃えることができる。

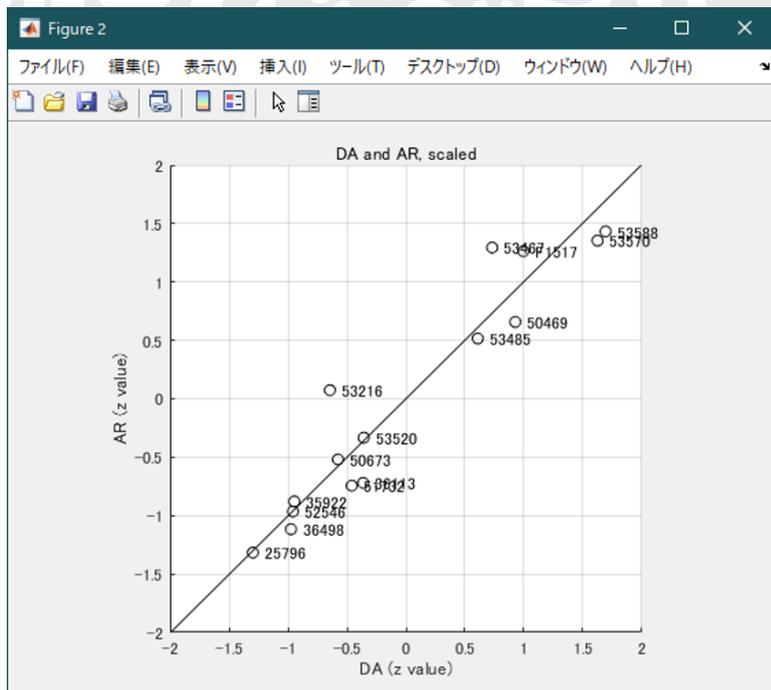


図4.ノーマライズ後のDAとARの散布図

図4はノーマライズを経た後の散布図である。ここでDAとARの相関係数 t を求め、 $1-t$ を相関係数に基づく距離として定義する。以上が相関係数に基づく距離の導出である。

<クラスター解析におけるユークリッド距離と相関係数に基づく距離>

- ・ユークリッド距離でクラスター解析を行う場合、径の大きさの近さで分類する。
- ・相関係数に基づく距離の場合、径の変化パターンの近さで分類する。

※図表は滋賀医科大学の松林潤先生作成のスライドより借用いたしました。