データサイエンス・AI基礎

学籍番号:2713240012-7

氏名：今村優斗

提出日：6月13日

　私は、ヒストグラムや散布図などのデータの可視化や単回帰分析などのデータ分析が社会の様々なところで役に立っていると考える。具体的な例として、医療分野の生活習慣病や未知のウイルスによる病気などの問題についてである。

　まず、データの可視化が医療分野の社会課題に対し有効な理由として、ある一つの要素に対して得られているデータを分類し比較することができるからである。例えば、ある生活習慣病における年齢層の人数を棒グラフや円グラフにすることによって、どの年齢が多く、どの年齢が少ないのか比較することができる。さらに、異なる年代でのグラフを比較することによって、昔と今ではどの割合がどのように変化したか調べることができる。これらから特定の要素によってデータを分けグラフにすることで、ある病気に対する因果関係を見つけることができるようになる。また、２変数以上のデータを可視化することができるモザイクプロットやヒートマップ、レーダーチャートによってさらに詳しい関係を得ることができる。例として、ある未知のウイルスの感染者数を全世界規模でヒートマップによってあらわすことで世界のどの地域で感染者が多く、どの地域が少ないのか可視化できる。これにより、感染源がどの地域であるかの推測や、そのウイルスが気候や湿度などの外的な影響を受けるものなのかどうかを判断する根拠となることができる。また、そのようなウイルスにかかった患者について、体温や血圧、心拍数、倦怠感、頭痛などの症状、白血球数などの検査による値を要素としてレーダーチャートを作る。それを他人のレーダーチャートと比較することで、そのウイルスがどのような特徴を持っているか発見できる。例えば、1000人程度の同一のウイルスに感染していると疑われている人たちのレーダーチャートを比較して全員の血中酸素濃度や呼吸の苦しさの項目での数値が高かった場合、そのウイルスは呼吸不全を起こすものであり、肺や赤血球に対して何らかの害があると絞ることができる。また、同じ人のレーダーチャートを日別に比較することによって、グラフの形の変化から様々なことを推測することができる。具体的には、グラフの形が小さくなっている場合、快方に向かっていることが分かり、グラフの形が変化して違う形に変化した場合、そのウイルスの性質が変化したと推測でき、現在の治療法が適切なものか判断したりできる。これらのことから、データの可視化を行うことで、医療分野において様々な情報を得ることができ、問題解決へとつながる。

　次に、単回帰分析などのデータ分析の活用について考える。単回帰分析を行うことによって、ある変数が変化することによって結果がどの程度変化するものかがわかる。例えば、生活習慣病の症状のひどさをyとして睡眠時間をxとし、データを集める。その後最小二乗法によって関数を作成すると、睡眠時間と生活習慣病の関係をあらわすことができる。この関数を利用すると、睡眠時間から生活習慣病にかかったときの症状の程度を推測することができる。また、未知のウイルスに対して重回帰分析を行うことでそのウイルスがどのような症状を持っているか調べることができる。例えば、ｙを重症度として、x1を血中酸素濃度、x2を体温、x3白血球数のように決定し、関数を最小二乗法によって算出する。そこから得た関数から、そのウイルスはどの要素が最も重症化につながっているのかを見抜くことができる。そうすることで、医療現場においてどの患者が重症化しやすいのか区別することができ結果的に多くの患者を効率よく治療することにつながる。このように、単回帰分析は得られたデータの中から関数を作成することによって、データ以外の人がどのような結果になるか推測することができる。

　これらのことから、データの可視化や分析が生活習慣病や未知のウイルスによる病気への対応を手助けしてくれることが分かった。データの可視化は、割合を比較することによって分けられた要素が関係しているものか判断することができ、直感的に因果関係を理解することができる。それに対して、データの分析は得られたデータから未知のデータがおおよそどの程度の結果に落ち着くか推測することができる。両者を上手に組み合わせることによって医療分野での社会問題の解決につながると考えた。