



データの可視化について

近年、第5世代移動通信システム（5G）やAIの台頭により、データの分析と活用が官民間問わず注目されています。しかし、データ分析には様々な注意が必要です。間違った分析により意思決定に悪影響がおよぶことは避けたいところです。

そこで、本稿では「国民生活基礎調査」での京都府の年齢階級別平均睡眠時間を例にとり、分析における最初のステップであるデータの可視化について注意点をご紹介致します。

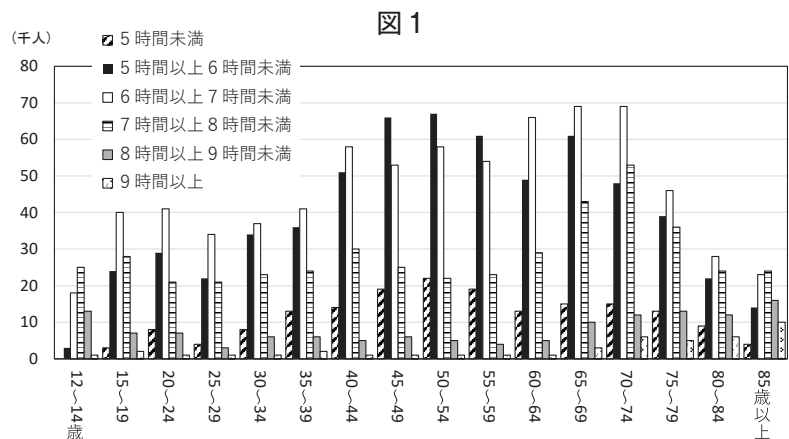
図1は京都府の平均睡眠時間の年齢階級別データをそのまま可視化したグラフです。一見、年齢階級別に睡眠時間ごとのグラフが作成されており、問題なさそうに思えます。しかし、このグラフは分析したい目的に合致しているでしょうか。

おそらく、こういったデータを分析する際に一般的に目的とするのは、「年齢に応じて睡眠時間は変化しているのかどうか」だと思います。上記のグラフは、人数を表しており、元のデータでは年齢階級ごとの人数が異なるため、年齢階級ごとの人数の違いが強調されてしまい、目的と合致しないグラフとなってしまっています。

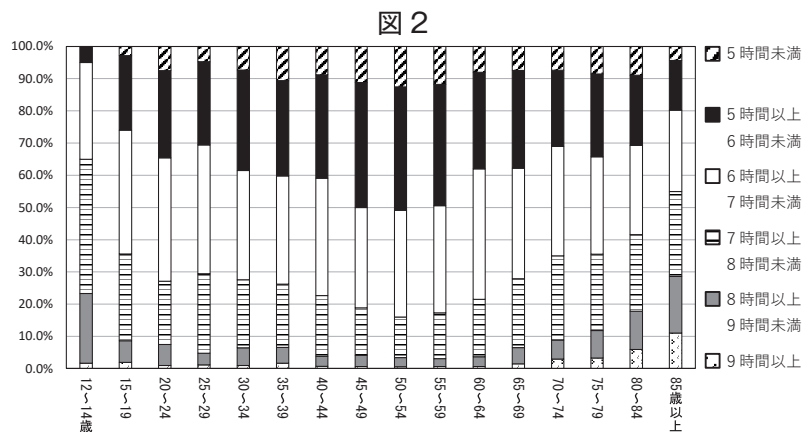
そこで、年齢階級ごとの人数の差をなくすため、年齢階級ごとに睡眠時間別の人数の割合をとり、集合棒グラフ化したものが図2になります。例えば「5時間以上6時間未満」と「7時間以上8時間未満」とを比較すると、20歳代以下や70歳代以上ぐらいでは7時間以上8時間未満の睡眠をとる方の割合が高いが、20歳代以上から65歳以下では、5時間以上6時間未満の睡眠時間の割合が大きくなってことが分かり、年齢別の睡眠時間の変化が把握できると思います。

このデータをもとに、例えば、年齢ごとの病気の発生率（生活習慣病やがん等）のデータなどを加えると睡眠と病気について興味深い分析が出来そうですね。

以上のように、漠然とデータを可視化するのではなく、このデータはどのようなデータか（今回でいえば、年齢ごとに総人数が異なる。）を把握した上で、必要な加工を行い適切な形式のグラフにすることは、データの特徴を把握する上で大変重要です。データを可視化する際は、「分析の目的」を第一に意識して、正しい見識が得られるよう心がけましょう。



資料 令和元年国民生活基礎調査（厚生労働省）
注 グラフは各年齢階級ごとの不詳分を除いて作成している。



資料 令和元年国民生活基礎調査（厚生労働省）
注 不詳分を除いた各年齢階級ごとの合計人数で、年齢階級ごとの各平均睡眠時間の