

2020年5月

COVID-19 のデータセットを取得する【第9回生物統計学】

1. はじめに

弊社では個別に統計解析のご依頼をいただいております。その中で、お客様が取得されたデータがデータセットとして機能していない場合が散見されています。そこで、今回は東京都の COVID-19 に関するデータセットを取得してその中身を見ながらデータセットについてみていこうというものです。

2. データセットの取得

データセットは、以下の URL から取得できます。

https://stopcovid19.metro.tokyo.lg.jp/data/130001_tokyo_covid19_patients.csv

データセットの形を観察してみると 1 行目に変数名が入力されていて、2 行目以降が実際のデータの様です。1 行につき 1 例のデータを記載していますが、このような入力形式がデータセット作成の基本的なポイントになります。5 列目に時系列データがありますので、<https://stopcovid19.metro.tokyo.lg.jp/> で公開されている「新規患者に関する報告件数の推移」のようなグラフを簡単に作成できますね。

取得したデータでは検査数が入力されているので東京都が公開している「検査実施人数（陰性確認を除く）と陽性率の推移」のグラフを作成することはできませんが、「検査実施件数」のデータは日付ごとにひろっていけば作成することができます。

COVID-19 のデータ以外にも都や国が収集したデータはインターネットを介して取得することができます。傷病者数や有訴者数も公開されているので機能性表示食品の開発に役立つデータが転がっている可能性もありますね。

3. まとめ

今回は COVID-19 のデータセットを調査してみました。今後、このデータを集計する方法を紹介しますので楽しみに(#^^#)

4. 参考文献

- 東京都 新型コロナウイルス感染症対策サイト (<https://stopcovid19.metro.tokyo.lg.jp/>)