2022年7月

度数分布の階級数を決定する「スタージェスの公式」【第109回生物統計学】

1　概要

臨床データの集計や解析においては、アウトカムをどのように表現するのかが重要になります。等間隔に区切られた数値で大小関係や差を計算できるデータを「間隔尺度」といい、体重や血液マーカーの数値など、臨床データとしてもなじみ深いものも含まれます。この感覚尺度のデータの広がりを表現するのに、度数分布図（ヒストグラム）や箱ひげ図を用いて示すことも多いです。

「度数分布」とは、収集したデータをいくつかの階級に分けた際、それぞれの階級にいくつのデータが所属しているかカウントしたデータの分布状況のことをいいます。度数分布を見る際には、ヒストグラムを一目見て分布の特徴が捉えられるように階級の数または各階級の幅を決定する必要があります。今回は、度数分布の階級数の決定に用いられる「スタージェスの公式」を紹介します。

2　スタージェスの公式の概要

スタージェスの公式とは、「データ数が n ある場合、階級数は 1+Log(2n) とするのがよい」というものです。手元にある標本データが正規分布をとると仮定した場合、データをいくつに区切れば正規分布に近いデータ数の分布が最も見えやすくなるか計算することができます。

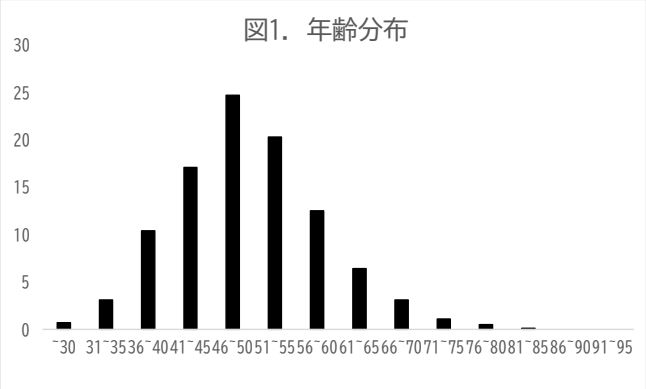
3　臨床データを用いたスタージェスの公式の活用例

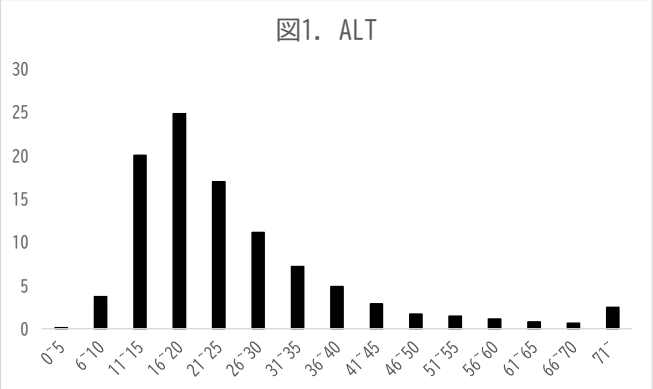
ある病院の人間ドック受診者 9,072 例の年齢分布を示す際、この例数をスタージェスの公式に当てはめると、階級数は 1 + Log(2*9,072) = 1 + 13.1 = 14.1 となり、およそ 14 階級に分けるのが目安になると思われます。これをヒストグラムに表すと図 1 のようになり、左右への歪みが少ない分布であり、正規分布と考えて検定してよい分布であると考えられます。

一方、11,870 例の血液検査データから得られた ALT 値の分布を当てはめると、階級数の目安は 1 + Log(2*11,870) = 1 + 13.5 = 14.5 でおよそ 15 階級となり、ヒストグラムに表すと図 2 のように右裾に長く尾を引く上に、最大値付近で頻度が再び上がっていることが分かります。この分布は図 1 に示した年齢のように正規分布に近いと見なすことができず、正規分布を仮定しない検定手法を用いることが適切であると考えられます。

以上のように、度数分布を視覚的に捉えるためには分布の特徴をよく表したヒストグラムを作成することが有益であり、そのためにはスタージェスの公式を用いて適切な階級数を設定する必要があることが分かります。

ORTHO MEDICO
Contact Information
ORTHOMEDICO inc., Planning Department
2F Sumitomo Fudosan Korakuen Bldg.,
1-4-1, Koishikawa, Bunkyo-ku, Tokyo, 112-0002 Japan
Phone: +81-3-3818-0610
E-mail: planning-department@orthomedico.jp





4　まとめ

統計解析に際しては、対象とするデータの分布を理解し、適切な手法を選択する必要があり、そのために視覚的に分布を捉えられるヒストグラムが有益になります。分布が見えやすいヒストグラムを作成するのに役立つ公式があることを知っておくとデータの理解に役立つことと思います。

5　参考文献

- 荒瀬康司. 論文投稿に際しての統計学的記述の留意点. 人間ドック. 2018; 33: 557-570.

ヒト臨床試験（ヒト試験）で得られる結果は、様々な誤差を含んでいます。この誤差を小さくすることで介入効果を増大させることができます。オルトメディコは、多分野の専門家を有するため、様々なアプローチにより誤差を最小化する試験運営が可能です。引き続き、皆様にご満足いただけるような高品質なヒト試験を提供させていただきますので、今後ともどうぞ宜しくお願い申し上げます。

Contact Information
ORTHOMEDICO inc., Planning Department
2F Sumitomo Fudosan Korakuen Bldg.,
1-4-1, Koishikawa, Bunkyo-ku, Tokyo, 112-0002 Japan
Phone: +81-3-3818-0610
E-mail: planning-department@orthomedico.jp