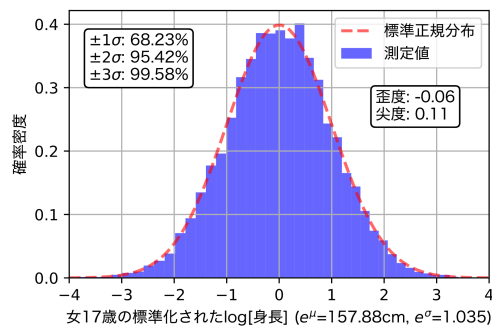
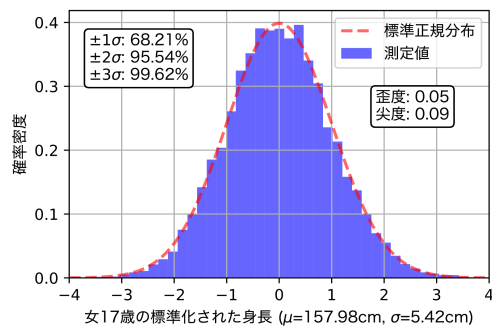
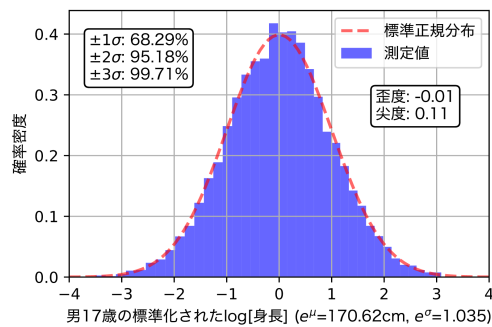
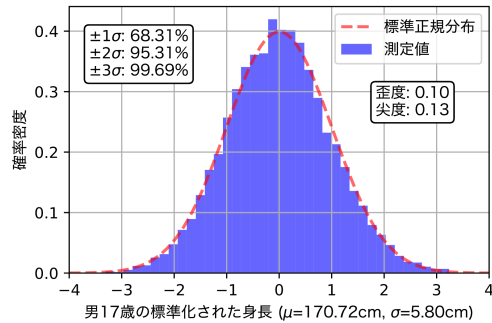


身長と体重の正規性

著者：関勝寿 公開日：2024 年 10 月 12 日 ソース：<https://sekika.github.io/2024/10/12/body/>

文科省の学校保健統計調査のデータについて、身長と体重が正規分布とどれだけ近いかを視覚化した。また、指定した年度のデータをダウンロードして、指定した年齢のデータについて、同様のグラフを表示するプログラムを公開する。なお、このプログラムは正規分布の必然性に関する考察の付録として作成したものであり、論文とあわせて読まれることを期待している。

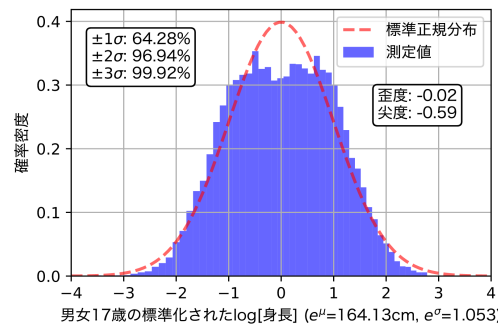
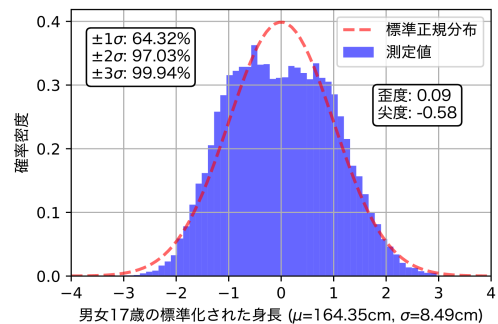
1. 身長



この図は、学校保健統計調査の2022年度の17歳の男女別の身長を、それぞれ正規分布と対数正規分布と比較したものである。

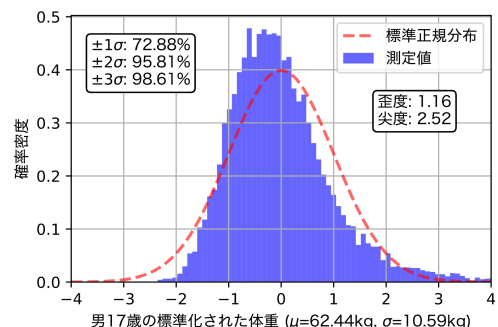
- ・正規分布については、身長を標準化してヒストグラムを描き、標準正規分布と比較した。横軸には平均 μ と標準偏差 σ の値を記入している。
- ・対数正規分布については、身長の対数を標準化して、標準正規分布と比較した。対数正規分布については、 e^μ と e^σ の値を記入している。横軸0の目盛りが e^μ cmの身長に相当し、目盛りが1増えるごとに身長が e^σ 倍となる。
- ・いずれの図にも、平均値から $\pm 1\sigma$, $\pm 2\sigma$, $\pm 3\sigma$ に収まるデータの割合を示している。また、歪度と尖度を示している。正規分布では歪度と尖度は0となる。

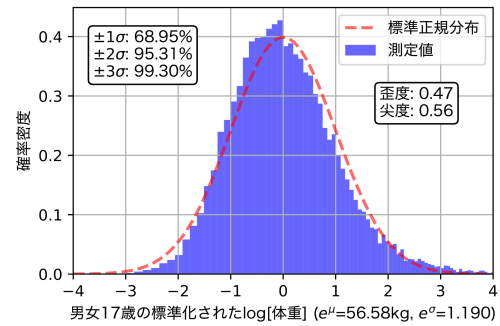
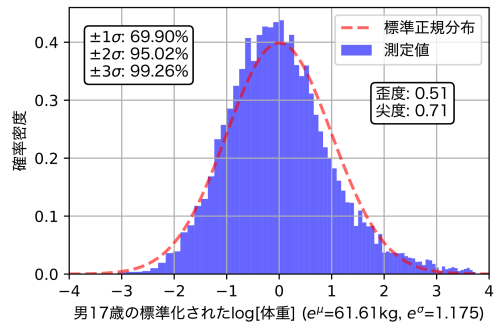
17歳の身長のデータは、男女ともに正規分布、対数正規分布いずれもよく適合した。歪度と尖度はいずれも絶対値が0.5以下であることも、正規分布および対数正規分布とよく適合していることを示している。



この図は、男女同数であると仮定して、男女合計のグラフを描いたものである。男の方が女よりも身長の平均が高いため、2つの正規分布が合わさった形となり、尖度が-0.58となって若干正規分布からずれた。17歳の身長平均は男女で13cmほどの差があり、これは2 σ に相当することから、-1と+1のあたりにそれぞれピークがあり、ピークの間の谷間が平坦化したようなグラフとなっている。

2. 体重





同様の解析を男女あわせて行った。やはり正規分布ではあわないが、対数正規分布ではほどに適合した。男女の平均体重差は10kg すなわち1σ程度しかないため、男女17歳の身長のように明確な2つのピークがあらわれなかった。むしろ、男女別に解析した場合と比べて、より対数正規分布に近くなった。

3. プログラム

このページの図を書くために作成したPythonのプログラムを掲載する。このプログラムは、学校保健統計調査の指定した年度のデータをダウンロードして、指定した年齢のデータについて、男、女、男女の身長、体重、それぞれを正規分布、対数積分布で標準化してグラフを描き、標準正規分布と重ねて表示する。

2006年度から2022年度までのデータ解析ができることが確認されているが、このプログラムはe-Statのデータ公開形式に依存しているため、今後e-Statにおけるデータ公開形式が変化した場合には使用不可となる。

実行手順

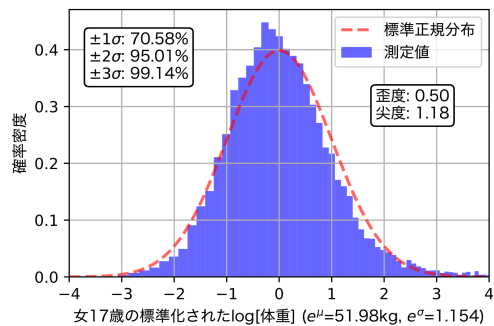
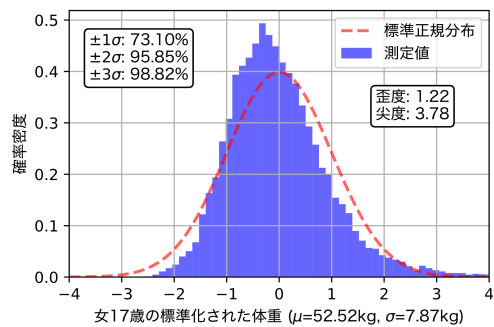
- [ここからダウンロードする](#)。
- `pip install beautifulsoup4 matplotlib numpy pandas Requests scipy xlrd openpyxl`
- 日本語のフォントを使うために、フォントファイルのパスをFONT_PATHに設定する。詳しくは[Matplotlibで日本語のグラフを作成する](#)を参照。
- Python 3 で実行する。コマンドラインオプションは以下の通り。

usage: body.py [-h] [-y YEAR] [-a AGE] [-e [EXT]] [-b] [-k]

学校保健統計調査の身長体重データ解析

optional arguments:

- h, --help show this help message and exit
- y YEAR, --year YEAR 解析する年 (西暦、2006 年以降、デフォルトは 2 年前)
- a AGE, --age AGE 年齢 (デフォルトは 17)
- e [EXT], --ext [EXT] 図のファイルの拡張子 (デフォルトは svg, -e のみで一覧表示)
- b, --black 白黒の図を生成
- k, --kolmogorov Kolmogorov-Smirnov 検定の p 値をグラフに表示



この図は、学校保健統計調査の2022年度の17歳の男女別の体重である。いずれもグラフは正規分布からはかなり外れ、歪度は1を超え、尖度は2を超えた。体重については平均よりも重い場合について、より平均から大きく乖離しやすいという非対称性があることから、歪度が生じる。そして、標準偏差が重い体重の人によって大きくなることから、平均の近くに多くのデータが集まることになり、尖度が大きくなる。

一方で、体重は対数正規分布にほどほどに適合していた。体重の対数を基準とすることで、歪度がより小さくなるだけでなく、尖度も小さくなった。このように、体重は正規分布よりも対数正規分布の方がよくあうということが確認できた。

