日本の子供の体重分布に関する統計的研究

2017SS083　坪井紗輝
指導教員：小藤俊幸

1　はじめに

子供の体重は，食生活の欧米化，テレビゲームやスマートフォンの普及による生活リズムの変化により，自分の親の世代と比べて変化があると考えられる．では実際，子供の体重にはどのような変化が起きているのかを文部科学省の学校保健統計調査のデータを用いて，それぞれの世代ごとの体重の変化を分析する．

学校保健調査は，統計法に基づく基幹統計調査として，明治 33 年に生徒児童身体検査統計として始まり，幼稚園，小学校，中学校，高等学校に在籍する満 5 歳から 4 月 1 日時点での 17 歳までの児童・生徒を対象に児童・生徒の発育・健康状態を明らかにすることを目的とした調査である [1].

2　対数正規分布

正の値をとる確率変数 X が対数正規分布に従うとは，$Y = \log X$ が正規分布に従うことをいう．X の分布関数を $F(x)$ とすると，

$$F(x) = P(X \le x) \;=\; P(\log X \le \log x)$$
$$= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\log x} e^{-\frac{(y-\mu)^2}{2\sigma^2}}\,dy \tag{1}$$

となることから，密度関数 $f(x)$ は

$$f(x) = F'(x)$$
$$= \frac{1}{\sqrt{2\pi}\sigma\, x}\, e^{-\frac{(\log x-\mu)^2}{2\sigma^2}} \tag{2}$$

で与えられる．

また，標準正規分布の分布関数

$$\Phi(u) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u} e^{-\frac{y^2}{2}}\,dy \tag{3}$$

を用いると，対数正規分布の分布関数は

$$F(x) = \Phi\left(\frac{\log x - \mu}{\sigma}\right) \tag{4}$$

と表される．あるデータの累積度数 g_i（g_i は x_i 以下のデータの度数）が

$$g_i = \Phi\left(\frac{\log x_i - \mu}{\sigma}\right) \tag{5}$$

で与えられるとすると，両辺を Φ^{-1} で変換することにより，

$$\log x_i = \sigma\,\Phi^{-1}(g_i) + \mu \tag{6}$$

が得られる．したがって，あるデータが対数正規分布に従うならば，横軸を累積度数，縦軸を log (階級値の右端) にとってデータを表示すると，直線となるはずである．ここでは，この表示を対数正規 Q-Q プロットと呼び，データが対数正規分布に従うかどうかの判定に用いることにする [2].

3　体重の比較

体重は青年期までは対数正規分布に従うが，青年期を超えると体重は個人差が生まれ対数正規分布に従わなくなると言われている [3]．体重は，肥満のグループと普通のグループの 2 つに分かれて二重対数正規分布の可能性が考えられている [3]．今回は，どの年代でも体重は対数正規分布に従うのか，また，2 つのグループに別れる二重対数正規分布となるか検証する．昭和 55 年の体重のデータ，平成 10 年の体重のデータをサンプリングし，得た分布を比較する．今回は 12 歳男子，17 歳男子を検証した．

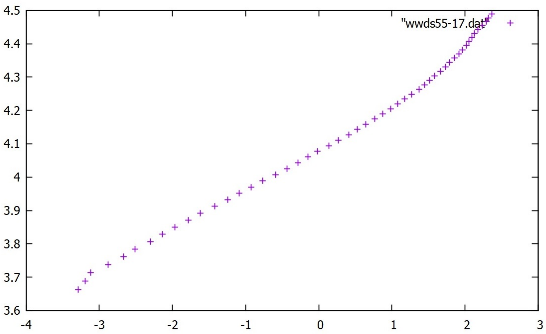

図 1　昭和 55 年 17 歳男子 Q-Q プロット

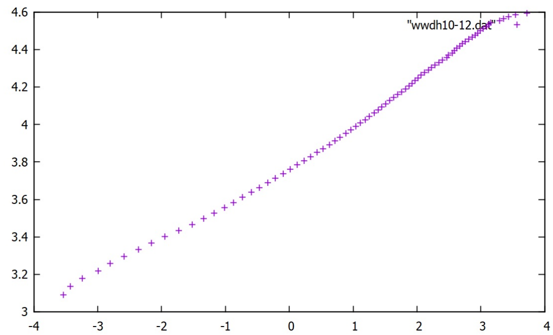

図 2　平成 10 年 12 歳男子 Q-Q プロット

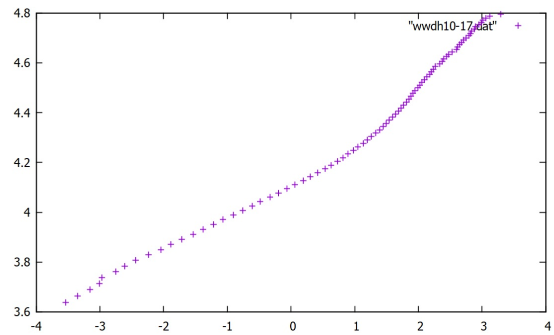

図 3　平成 10 年 17 歳男子 Q-Q プロット

検証した結果，昭和 55 年 12 歳男子，17 歳男子ともに体重の重い人は線から少し直線からずれているが，おおよそ一直線に並んだため対数正規分布に従っていると考えられる．また，平成 10 年 12 歳男子もおおよそ対数正規分布に従っていた．一方で，平成 10 年 17 歳男子の Q-Q プロットでは直線のグラフではなく折れ線グラフに近い形となり，このことからグラフは二極化していると推測される．

4　2 つのグループの比較

体重のデータを Q-Q プロットを用いて比較した結果，平成 10 年 17 歳男子は折れ線のようなグラフとなり，2 つのグループに分けられるのではないかと推測される．以下より 2 つのグループをグループ a，グループ b とする．2 つのグループがどのように分かれているかを判断するために，平成 10 年 17 歳男子のデータを Excel を用いてそれぞれの回帰直線から傾きと y 切片を求めた．C_1 は分散が小さいが平均が大きくなるグループ a の割合とし，C_2 は分散が大きいが平均が小さくなるグループ b の割合とする．

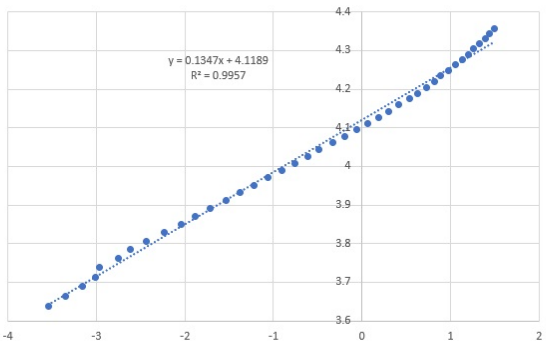

図 4　グループ a

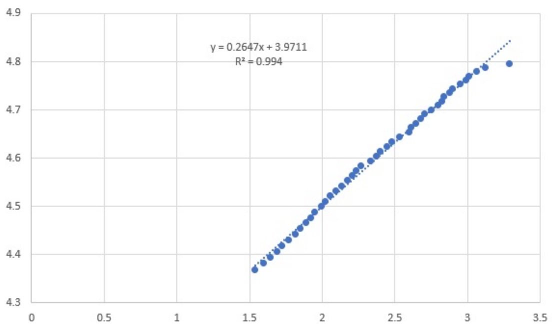

図 5　グループ b

その結果から，(2) を用いてそれぞれの密度関数を求めた．グループ a の密度関数を $f_1(x)$，グループ b の密度関数を $f_2(x)$ とすると，全体の密度関数は，

$$f(x) = C_1 f_1(x) + C_2 f_2(x) \text{ と表される．}$$

ヒストグラムと上手く重なり合う密度関数を得るため，C_1 と C_2 を調整して求めた結果，グループ a の割合は 94 %，グループ b の割合は 6 %となった．

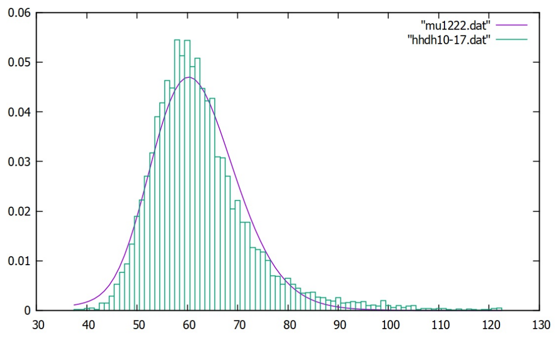

図 6　平成 10 年 17 歳男子のヒストグラムと密度関数

5　考察

参考文献より，体重の重いグループと体重の普通のグループに別れ，それら 2 つが二重対数正規分布となる要因であると推測していた ([3],p.110)．しかし，Q-Q プロットで 2 つのグループに分かれることは推測できたが，その 2 つのグループが個人の体重の重さでなく分散が大きいグループと小さいグループに別れることが分かった．この結果は，体重は身長とは違い，自分自身で操作がしやすく複雑なためであると考えられる．特に最近では，インターネットの普及により，ダイエットや痩せていることが魅力的であるような情報が増えたことで，痩せ願望を持つ子供が増え，体重の重いグループにいた人が短期間で軽いグループに移行する．さらに近年のテレビゲーム, スマートフォンの普及により外に出て遊ぶことよりも家に居る事が増えた事，そして日本の食卓に和食だけでなく，洋食も取り入れられるようになり, カロリーを多く含むものを食べるようになったため肥満児の割合が増えた．そのため軽いグループから重いグループに移行した人もおり，このように個人の努力や痩せやすい，痩せにくい遺伝等の様々な要因が絡み合って 2 つグループに分けられたと推測する．

6　おわりに

子供の体重は必ずしも対数正規分布の従うとは言えず，生活習慣等で個人差が出ると推測できた．さらに体重の分布は裾の広がりが大きく分散が大きいグループと分散が小さいグループに分けられ，体重の重いグループと普通のグループに別れる事が二重対数正規分布の要因と判断出来なかった．今後の課題は二重対数正規分布となる要因を研究することである．

参考文献

[1] 文部科学省：「学校保健統計調査－調査の概要」．`http://www.mext.go.jp/b_menu/toukei/chousa05/hoken/gaiyou/chousa/1268648.htm` (最終閲覧日：2021 年 1 月 1 日)

[2] 小藤俊幸：「考える力をつけるための微積分教科書」．学術図書出版，東京，2019.

[3] 松下貢：「統計分布を知れば世界が分かる」．中央公論新社，東京，2019.