

品質管理と数学③

～ヒストグラムと箱ひげ図～

北海道白老東高等学校
教諭 工藤大輔

1 ヒストグラム・・・何に使うの？

現在の数学Bや数学基礎の教科書もそうであるが、統計を学ぶ上で、始めにヒストグラムを扱う。しかしながら、データ全体の傾向を調べることや、相対度数により全体に占める割合を調べたりすることにとどまっている。

一方、品質管理では、同一規格の製品等に対し、ヒストグラムを作成するため、ヒストグラムから製品の傾向や不良品の原因を読み取り、対策を練る等、ヒストグラムは非常に重要なものとなっている。

授業ではヒストグラムの作成までの流れを扱い、実際に作成した上で、どのように見るかを扱った。また、ヒストグラムの授業の後、「箱ひげ図」についても扱い、作成時の両者の違いや見方を扱った。

2 ヒストグラムの作成で

(1) 階級数の決め方

階級数については、様々な方法があるが、目安として用いる。

ア データ数の正の平方根を用いる方法

データの総数を n 、階級数を k とすると、

$$k = \sqrt{n}$$

品質管理関係の多くの文献をみても、この式で階級数を決めているものが多い。一般の電卓でも計算することができるため、算出が容易であると考えられる。

イ 「ステージス(Sturges)の公式」による方法

データ数を n 、階級数を k とすると、

$$k = 10 \frac{\log_{10} n}{\log_{10} 2} \times 10 \log_2 n$$

教科書等の階級数の目安もこの公式によるものであるが、本校では数学Ⅱが選択科目であり、数学基礎の受講者の半分以上が数学Ⅱを履修していない現状があり、紹介程度にとどめることとした。

ウ 各自定める方法

階級の幅を予め決めておき、そこから階級数が決まる方法である。今回はコンピュータを使わずにヒストグラムの作成を行ったため、この方法で授業を行った。

(2) 階級の幅

階級の幅は次の通り決められるが、授業では紹介程度にとどめた。

階級数を k とすると、

$$\text{階級の幅} = \frac{(\text{データの最大値}) - (\text{データの最小値})}{k}$$

なお、階級の幅は、「～以上・・・未満」ということにしたのだが、以上・未満の意味を忘れていた生徒がほとんどで、確認をしながら進めていった。

3 ヒストグラムの書き方

ヒストグラムが階級の幅に含まれる度数とでつくられるグラフであることを理解させたいところである。つまり、度数は階級の幅に含まれるデータの量であるため棒グラフであることと、全体の傾向を調べるためのものであるから、棒の間隔を詰めて書くというところがポイントとなる。

(1) 度数分布表の作り方

度数分布表の作成までで、1時間をつかった。もれなく、正確に数え上げる方法で、「正」の字を使うことが一般的となっているが、国や文化が違えば方法も変わることも同時に触れることができた。

表1 画線法の例

日本、中国、韓国 等	ヨーロッパ、北アメリカ	南アメリカ
一 下 下 正		┌┐┐┐
「正」の正しい書き順を知らずに間違って書くと、他者が勘違いすることがある。	欧米の他に、東南アジアの一部の国も、これを使う。	アルゼンチン、ブラジル、チリなど。

表2 B高校3年生40名の100m走の記録

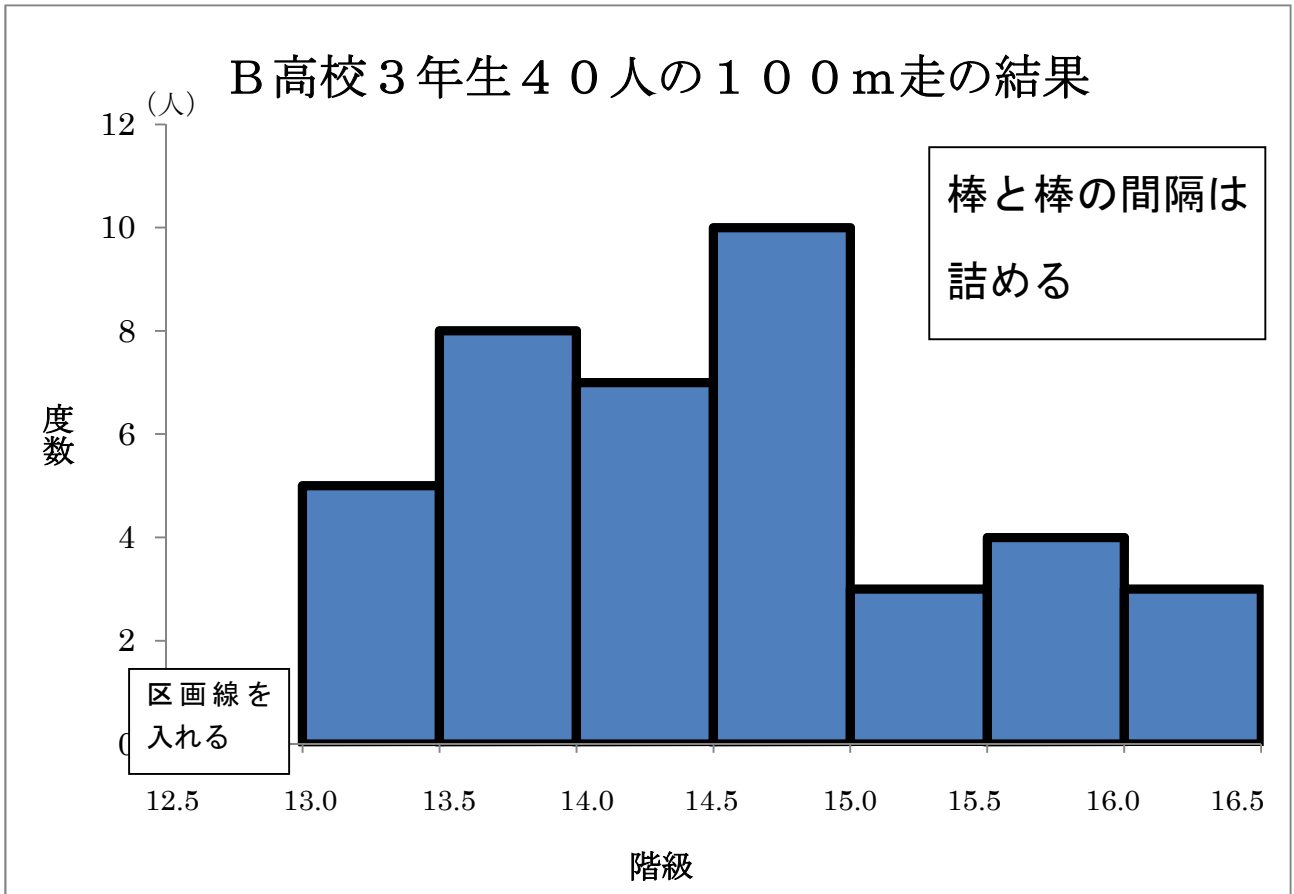
(2) 度数分布表の作成

数え上げることでさえできれば、あとは度数分布表を作成するだけであるが、度数分布表にも相対度数や階級値を求める必要があるのかということも視野に入れておかなければならない。

授業では数学基礎の教科書で掲載されているものを使用し、2つの高校の高校3年生の100m走の記録で、度数分布表を作成した。

階級 (秒)	度数 (人)	相対 度数 (%)	累積相 対度数 (%)
12.5 以上 13.0 未満	0	0	0
13.0 以上 13.5 未満	5	12.5	12.5
13.5 以上 14.0 未満	8	20.0	32.5
14.0 以上 14.5 未満	7	17.5	50.0
14.5 以上 15.0 未満	10	25.0	75.0
15.0 以上 15.5 未満	3	7.5	82.5
15.5 以上 16.0 未満	4	10.0	92.5
16.0 以上 16.5 未満	3	7.5	100
合計	40	100	

(3) ヒストグラム作成で

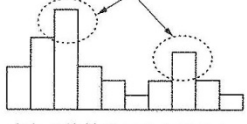
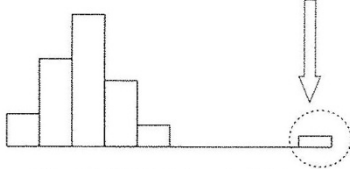


(4) ヒストグラムの見方

品質管理では、そのグラフの山を見ることで、全体の傾向を知ることの他に、原因を分析したり、今後の方向性を読むことをしていく。

表3 ヒストグラムの代表的なパターン

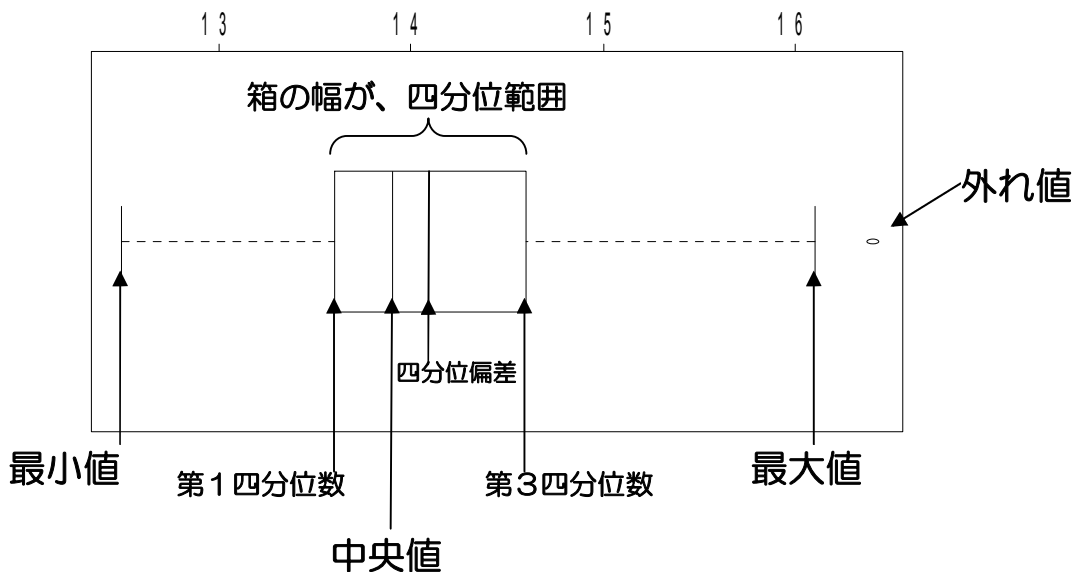
<p>①単峰性で左右対称</p> <p>データが集中している部分(峰、山)が1つ</p> <p>(A) 単峰で左右対称なヒストグラム</p>	<ul style="list-style-type: none"> ・ヒストグラムの基本となる形。 ・ヒストグラムを作成した際、一般的に表れる形。
<p>②非対称な形</p> <p>右方向に裾をひく分布 (右に歪んだ分布)</p> <p>左方向に裾をひく分布 (左に歪んだ分布)</p>	<ul style="list-style-type: none"> ・裾が広がっている部分が、山を形成している部分とは本質的に異なる、何か偏ったデータであることを指す。 ・左右のどちら側に裾を引くかを把握する必要がある。 <ul style="list-style-type: none"> ●右に裾を引く→右に(正の方向に)ゆがんだ分布 ●左に裾を引く→左に(負の方向に)ゆがんだ分布 ・この場合、理論的、または規格値などで下限(上限)が押さえられている場合に出る。

<p>③多峰性のヒストグラム</p> <p>データが集中している部分(峰, 山)が2つ</p>  <p>(B) 2峰性のヒストグラム</p>	<ul style="list-style-type: none"> ・山の数だけ、異質な集団が混在していることを指す。 ・この場合は、測定者ごと、測定機器毎に分けてヒストグラムを作成してみる。
<p>④外れ値がある</p> <p>データの大部分が含まれるデータ値の区間からかけ離れたところに位置するデータ</p>  <p>(D) 外れ値があるヒストグラム</p>	<ul style="list-style-type: none"> ・対象データの中に、明らかに異質なデータが少数個含まれていることを意味する。 ・このようなデータを「外れ値」と言う。外れ値は集団の平均値の算出等で影響を与えることが多い。 ・外れ値の存在は何らかの重要な「発見」に結びつく場合がある。

4 箱ひげ図

(1) 箱ひげ図とヒストグラム

箱ひげ図は、中央値や最大値・最小値、四分位数で図を作成する。四分位範囲により外れ値の基準を算出する。また、中央値を用いることで、外れ値の影響を受けやすい平均値よりも正確な分析が可能である。また、箱ひげ図からヒストグラムの概形を予想することも可能である。



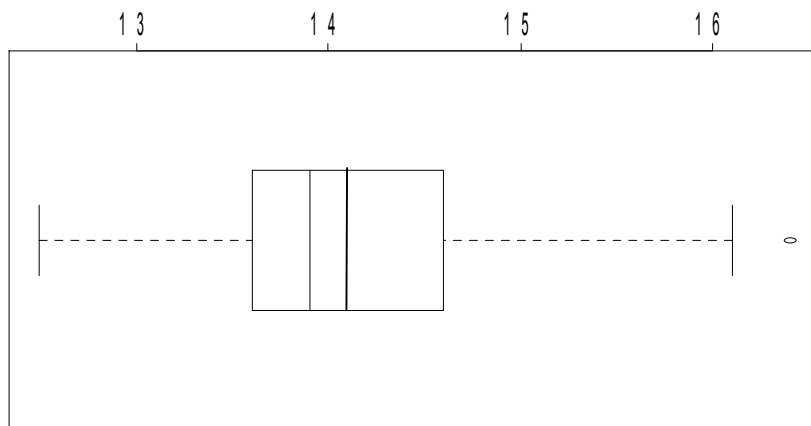
(2) 外れ値について

軽度の外れ値： $(\text{四分位範囲}) \times 1.5$ とし、これを超える場合は箱ひげ図では○で記入する。

極端な外れ値： $(\text{四分位範囲}) \times 3$ とし、これを超える場合は箱ひげ図では×で記入することとする。

(3) 箱ひげ図の見方

- ①箱の位置やひげの長さから、左右（縦図の場合は上下）が対称かどうかをチェックする。
- ②中央値から最大値（もしくは最小値）、までの距離が短ければ、ヒストグラムの山がその部分に寄っている。
- ③四分位範囲は箱の幅を示し、四分位範囲÷2の値が四分位偏差。これと中央値とのズレを確認することで、山の歪みをはっきりする。



A高校40人の100m走の結果

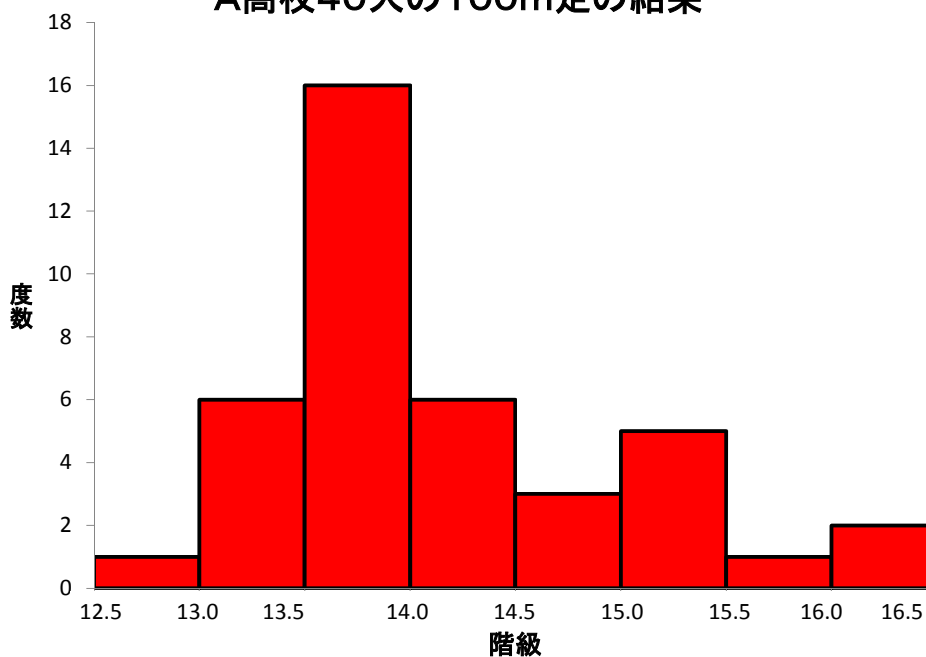


図1 箱ひげ図とヒストグラム

図1を見ると、箱ひげ図の示すとおりに、箱が左に寄っているため、ヒストグラムの山が左に寄っている。また、四分位偏差と中央値の間に差があるため、ヒストグラムの14.0以降の山に少し歪みがあることがわかる。