

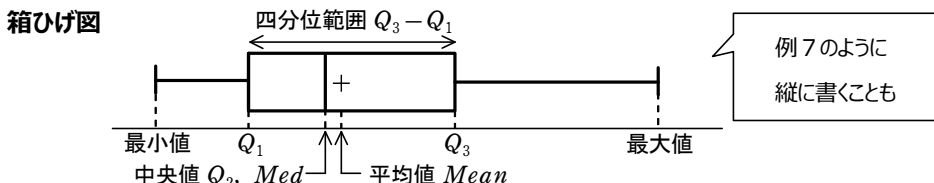
1 学習内容の説明 ⇒ 2 問題演習 ⇒ 3 振り返り（確認テスト・相互採点・リフレクションの記入）

【内容目標】箱ひげ図をかくこともでき、情報を読み取れるようになろう

□箱ひげ図

データの分布を見るための図に **箱ひげ図** と呼ばれるものがある。

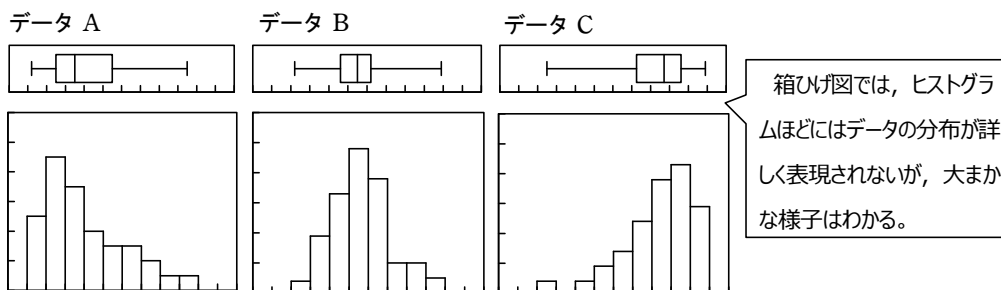
箱ひげ図は、データの最小値、第1四分位数、中央値、第3四分位数、最大値を、箱と線（ひげ鬚）で表現する図である。箱の長さは四分位範囲を表す。なお、箱ひげ図に平均値を記入することもある。



箱ひげ図は複数のデータの分布を比較するときに便利な図である。

□箱ひげ図とヒストグラム

下の図は、あるデータ A, B, C のヒストグラムと箱ひげ図との対応である。

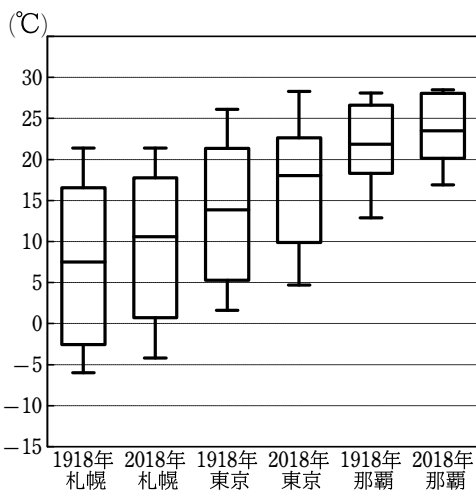


例7) 箱ひげ図によるデータの分析

右の図は、1918年と2018年の札幌、東京、那覇における、月ごとの平均気温のデータを箱ひげ図に表したものである。

この図から、たとえば、次のようなことが読み取れる。

1. 那覇は1年を通して暖かく、寒暖の差が小さい。
札幌は年間の寒暖の差が大きい。
2. どの都市も、1918年より2018年の方が気温が高めの傾向にある。
(特に Q_1 の上昇が著しい)



(気象庁ホームページより作成)

□外れ値

データの中に、他の値から極端に離れた値が含まれることがある。

そのような値を **外れ値** という。

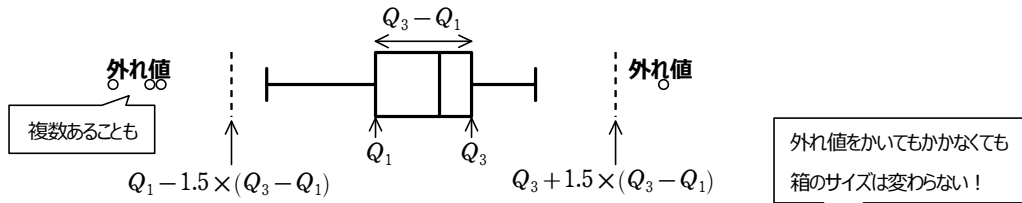
外れ値の基準は複数あるが、たとえば、次のような値を外れ値とする。

(第1四分位数-1.5×四分位範囲) 以下の値

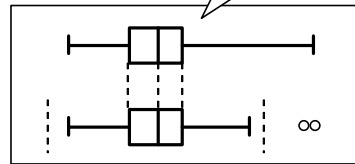
(第3四分位数+1.5×四分位範囲) 以上の値

外れ値がある場合、次の図のような箱ひげ図が用いられることがある。

外れ値は○で示している。また、箱ひげ図の左右のひげは、データから外れ値を除いたときの最小値または最大値まで引いている。

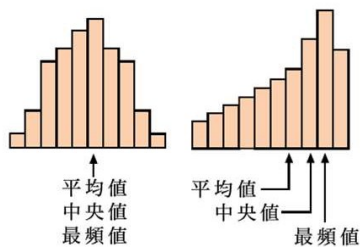


<注意> 外れ値を○で示す箱ひげ図をかく場合でも、四分位数は外れ値を除かないすべてのデータの四分位数であり、その値にもとづいて箱をかく。



外れ値は、測定ミスや入力ミスによる異常な値であることもある。一方で、外れ値の背景を探ることにより、問題発見があったり、問題解決の手がかりが得られたりすることもある。たとえば、販売員の販売成績を調べたとき、並外れて成績がよい販売員がいたら、その販売員の工夫を探ることで、全体の販売成績を上げる対策を見いだせる可能性がある。

たとえば、ある5人のテスト結果が15点、13点、12点、16点、98点のとき、平均値は30.8点、中央値は15点である。98点はその値より著しく高いため、98点の値を「はずれ値」として除いて考えると、平均値は14点と大きく変化するが、中央値は14点とあまり変化しない。このことから、平均値の方が中央値よりも「はずれ値」からの影響を受けやすいといえる。

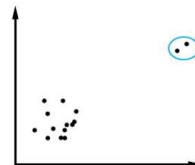


また、平均値、中央値、最頻値は、右の図のように、データが1つの山でほぼ左右対称に分布しているとき、互いに近い値となる。

一方、データが1つの山で左右対称に分布していないときは、平均値より中央値や最頻値の方が、データの代表的な位置を表すのに適切である。

さらに、相関係数も「はずれ値」の影響を受けやすい。

右の散布図で表されたデータの相関係数は0.885であるが、円の中の2つの「はずれ値」を除いて相関係数を求めると0.023となる。



このため、相関係数を考えるときは相関係数だけでなく、散布図も調べるのが重要である。