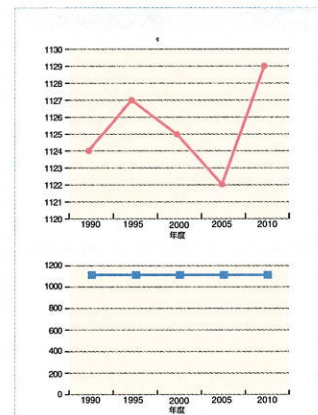
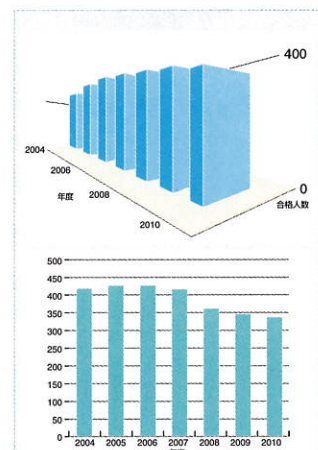


誤解を与えるグラフ



上のグラフでは、目盛りが0からはじまっていない。下のグラフに比べ、差が大きく見える。



上の棒グラフでは、数が増えているように見える。

グラフ化

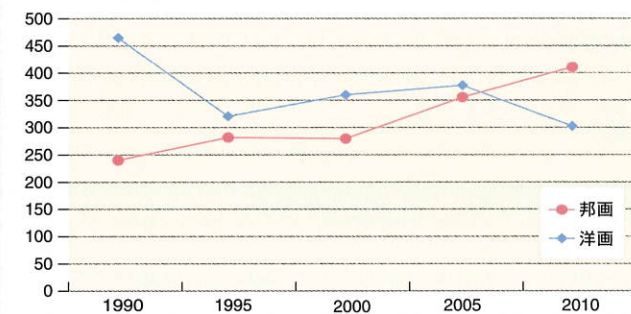
収集した数値データは、グラフに表現することで、データの特性を見つけやすくなる。

次のデータは、国内の邦画・洋画の公開本数である。

	1990	1995	2000	2005	2010
邦画	239	289	282	356	408
洋画	465	321	362	375	308

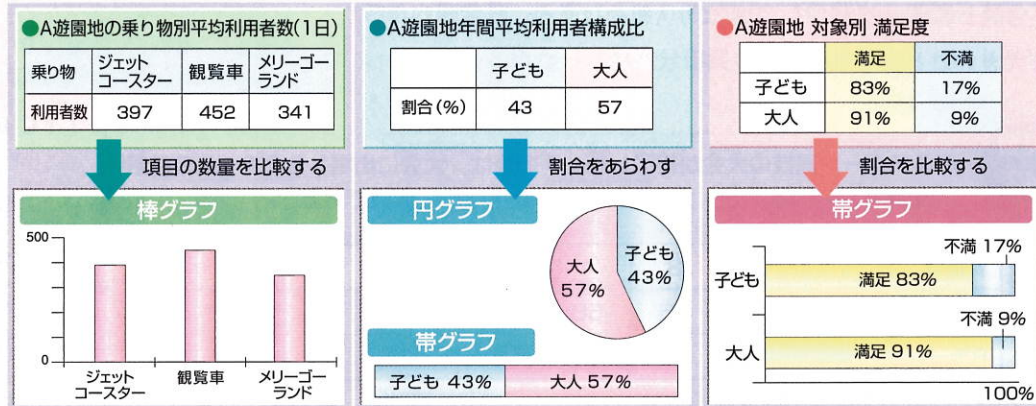
一般社団法人日本映画製作者連盟統計資料より

このままでは情報が読み取りにくいので、折れ線グラフにする。折れ線グラフは、時間の変化による数量の推移などをあらわすのに利用される。



このように、視覚化することで情報が読み取りやすくなる。グラフには、円グラフや棒グラフなど、さまざまな種類がある。グラフを作成する際には、「どのような数値データをグラフにするのか」、「何を表現したいのか」によって適切なグラフを選択する。また、グラフのつくり方によって印象が異なることがあるので注意が必要である。

さまざまなグラフ



実習 43

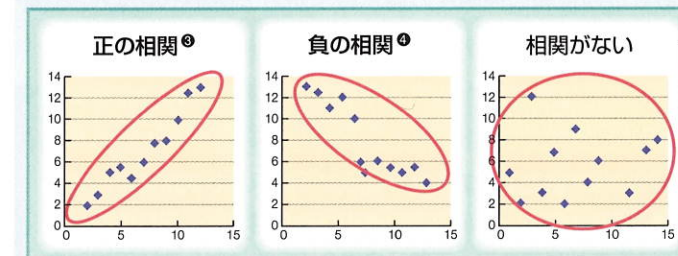
日本国内における邦画・洋画の公開本数の割合の推移がわかりやすいグラフを作成しよう。

PRACTICE

相関

二つのデータの関係を調べるには、データの相関^①を調べるとよい。相関を図であらわしたものが散布図^②である。

相関を示した散布図の例



分析の観点

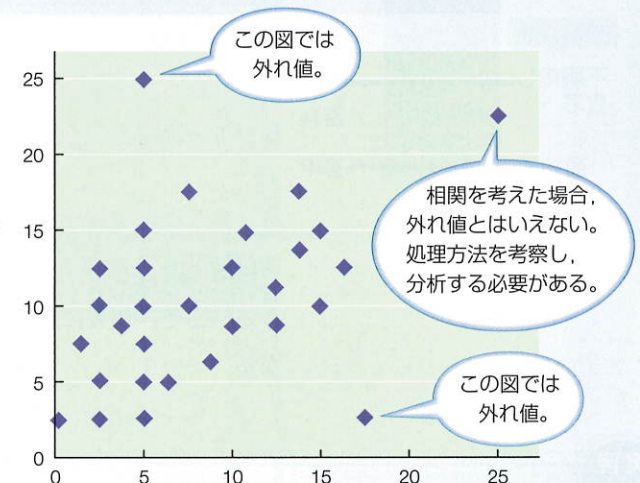
- 正の相関, 負の相関, 相関なし
- 直線, 曲線, 直線+曲線, その他

駅ごとの放置自転車の数と飲食店数を調べたところ、正の相関があったとする。相関関係を調べ、分析するには、次の2点について注意しなければならない。

1. 相関関係があるからといって、因果関係^③があるとはいえない。この例でいえば、「飲食店が多くなれば、放置自転車が増える」とは直ちにいい切れないということである。
2. ほかの変数を媒介とした擬似相関^④の可能性もある。駅前の通行人の多さという変数によって、飲食店の数と放置自転車の数に因果関係があるように見えているだけかもしれない。

外れ値(outlier)の扱い

ほかのデータの傾向から大きく離れた値を外れ値という。データを分析する際は、このような値を除外して扱うことが多い。ただし、外れ値はつねに異常な値であるとは限らない。転記ミスや入力ミスによる場合もあるが、そうでない場合には外れ値が発生した原因をよく考察し、分析する必要がある。



① 相関

二つの変数の間に関係があること。また、二つの変数の関係をあらわす指標を相関係数という。1に近い値であれば正の相関、-1に近い値であれば負の相関がある。0に近づくとき散布図にまとまりがなくなり、二つの変数に関係が見られないため相関がない。

② 散布図

二つの変数がどのような関係をもっているかを示した図。

③ 正の相関

一方の変数が増加すれば他方も増加するという傾向があるとき、二つの変数の間には正の相関があるという。

④ 負の相関

一方の変数が増加すれば他方は減少するという傾向があるとき、負の相関があるという。

⑤ 因果関係

原因と結果のつながりのこと。Aが原因でBという結果が発生した場合、AとBは因果関係があるという。

⑥ 擬似相関

因果関係がないのに因果関係があるように見えること。