

天候データの分析課題（統計処理編）

天候データを用いて、基本的な統計処理とグラフ化、検定について学習する。

データの概要

配布データは札幌、東京、福岡の3都市について2012年～2014年の天候データである。

[気象庁のサイト](#)からダウンロードし、「昼の天気」と「夜の天気」を分析用に調整している。

- ファイル名：weather_2012_2014.csv
- データの構造
 - 都市（要素：札幌、東京、福岡）
 - 日付（期間：2012年1月1日～2014年12月31日まで）
 - 平均気温（単位：℃）
 - 最高気温（単位：℃）
 - 最低気温（単位：℃）
 - 降水量（単位：mm）
 - 日照時間（単位：時間）
 - 降雪量（単位：cm）
 - 平均風速（単位：）
 - 昼の天気（要素：晴、曇、雨、雪）
 - 夜の天気（要素：晴、曇、雨、雪）

統計処理に関する設問

下記の設問について解答し、レポートにまとめてPDFで提出せよ。

分からない設問は飛ばしてもよい。

2014年のデータを用いて、下記の設問に答えよ。

下記の設問は、頻度データを扱う練習である。

1. 各都市の「昼の天気」と「夜の天気」について、各要素の出現頻度（度数）を表にまとめよ。
2. 各都市の「昼の天気」と「夜の天気」について、各要素の出現割合を割合の積み上げ棒グラフとして書け。
3. 作成したグラフを比較して考察せよ。

2012年～2014年のデータを用いて、下記の設問に答えよ。

下記の設問は、「頻度データの組み合わせ」と「確率」に関するスキルを確認するものである。

1. 各都市について「昼の天気」と「夜の天気」に関して全ての組み合わせ（「晴から曇」や「雨から雪」など）の出現確率を計算せよ。
2. 各都市について、「昼の天気」が"晴"の場合、「夜の天気」も"晴"になる確率が最も高い都市とその都市における確率を計算せよ。

下記の設問は、「数値要約」と「分布と傾向の把握」に関するスキルを確認するものである。

1. 各都市について、「平均気温」の最小値、25%点、中央値、平均値、75%点、最大値をそれぞれ求めよ。
2. 各都市の平均気温の推移を比較するため、横軸を日付、縦軸を平均気温として**わかりやすく**グラフ化せよ
3. 各都市の平均気温の分布を比較するため、横軸を都市、縦軸を平均気温とした「ボックスプロット」と「ヴァイオリンプロット」を作成せよ。
4. 上記の設問で作成した2つのグラフを比較し、作図手法としての特徴の違いを述べよ。

下記の設問は、「統計的検定」に関するスキルを確認するものである。

1. 東京と福岡の「平均気温」について、分布を確認するためヒストグラムを作成せよ。
2. ヒストグラムを確認し、データに「正規性」があるかどうか考察せよ。
3. 東京と福岡の「平均気温」の年間平均について、有意な差があると言えるかどうか**適切と思われる手法**で検定せよ。
4. 上記の設問で用いた手法が適切であると判断した理由を述べよ。
5. 検定結果について、本当に結果が正しいと思えるかどうか考察せよ。