

2-10

総合演習

本節では、第2章で学習した内容の総合演習を行います。

2-10-1 データのダウンロードとデータクリーニング 1

Step 1 作業用フォルダーの作成

本節で作成するファイルを保存するための「2-10」フォルダーを、2-1-1で作成した「データサイエンス」フォルダーの中に作成しましょう。

Step 2 データのダウンロード

気象庁のWebサイト「<https://www.jma.go.jp>」から以下のデータをダウンロードしましょう。

- 地点：東京・羽田・札幌・那覇・昭和（南極）
- 項目：日別値・日最低気温
- 期間：2016年1月1日～2020年12月31日

Step 3 ファイルの移動とファイル名の変更

Step 2でダウンロードした「data.csv」ファイルを、「ダウンロード」フォルダーから、Step 1で作成した「2-10」フォルダーに移動しましょう。次に、ファイル名を「5地点日別最低気温.csv」に変更しましょう。

Step 4 ファイル形式の変換とシート名の変更

「5地点日別最低気温.csv」ファイルをExcelブックファイル形式に変換し、「2-10.xlsx」というファイル名で保存しましょう。さらに、シートの名前を「5地点日別最低気温」から、「2-10-1」に変更しましょう。

Step 5 ウィンドウ枠の固定とデータクリーニング

B6でウィンドウ枠を固定し、データクリーニングをして図のようになります。但し、ダウンロードした時刻は揃える必要はありません。

	A	B	C	D	E	F	G
1	ダウンロードした時刻：2022/07/22 07:08:41						
2	5地点の日別最低気温 (°C)						
3							
4	年月日	東京	羽田	札幌	那覇	南極	
5	2016/01/01	3.6	5.9	-2.6	15.2	-4.4	
6	2016/01/02	2.6	3.9	-0.7	14.8	-4.5	
7	2016/01/03	3.0	4.8	-1.8	19.5	-5.3	
8	2016/01/04	4.4	5.1	-3.8	19.1	-7.3	
9	2016/01/05	5.1	8.4	-5.4	19.0	-4.5	
10	2016/01/06	7.1	7.9	-4.0	18.3	-4.7	

注意 南極の2017年8月12日のセルが空白になっています。このように、記録されていない値を欠損値と呼び、欠損値を含むデータの扱い方には様々な手法があります。この総合演習では、欠損値を含む2016年1月1日から2020年12月31日までの範囲を選択する手法を用いて、処理を進めましょう。

2-10-2 基本統計量

Step 1 シートのコピーとシート名の変更

「2-10-1」シートをコピーし、「2-10-2」シートを作成しましょう。

Step 2 基本統計量の算出

「2-10-2」シートにおいて、行を追加して図のように5地点5年分の日別最低気温の基本統計量を求めましょう。

	A	B	C	D	E	F	G
1	ダウンロードした時刻：2022/07/22 07:08:41						
2	5地点の日別最低気温 (°C)						
3							
4	基本統計量						
5	データの個数	1827	1827	1827	1827	1826	
6	平均	12.7	13.9	6.1	21.6	-13.2	
7	標準偏差	8.2	7.9	9.4	4.8	8.3	
8	最大値	28.5	29.0	27.4	29.7	1.4	
9	中央値	12.9	14.1	6.2	22.0	-12.0	
10	最小値	-4.0	-2.4	-14.9	6.1	-37.5	
11							
12	年月日	東京	羽田	札幌	那覇	南極	
13	2016/01/01	3.6	5.9	-2.6	15.2	-4.4	
14	2016/01/02	2.6	3.9	-0.7	14.8	-4.5	
15	2016/01/03	3.0	4.8	-1.8	19.5	-5.3	
16	2016/01/04	4.4	5.1	-3.8	19.1	-7.3	
17	2016/01/05	5.1	8.4	-5.4	19.0	-4.5	

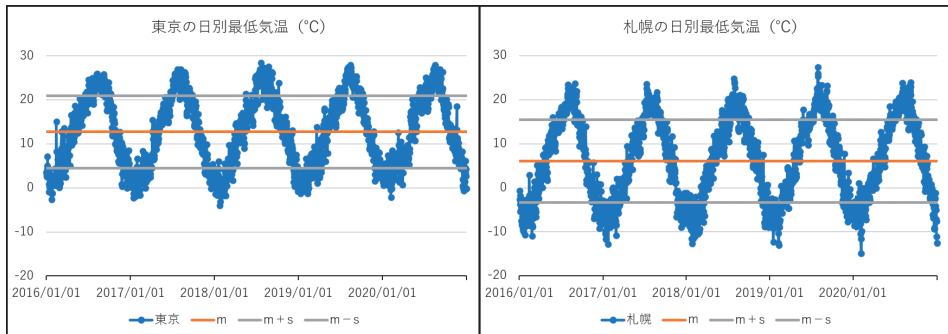
2-10-3 時系列データ・平均・標準偏差

Step 1 シートのコピーとシート名の変更

「2-10-2」シートをコピーし、「2-10-3」シートを作成しましょう。

Step 2 時系列データ・平均・標準偏差の可視化

「2-10-3」シートにおいて、東京と札幌の時系列データ・平均・標準偏差を可視化して比較するグラフを図のように作成しましょう。なお、平均をm、標準偏差をsで表記しています。



2-10-4 度数分布表とヒストグラム

Step 1 シートのコピーとシート名の変更

「2-10-1」シートをコピーし、「2-10-4」シートを作成しましょう。

Step 2 度数分布表の作成

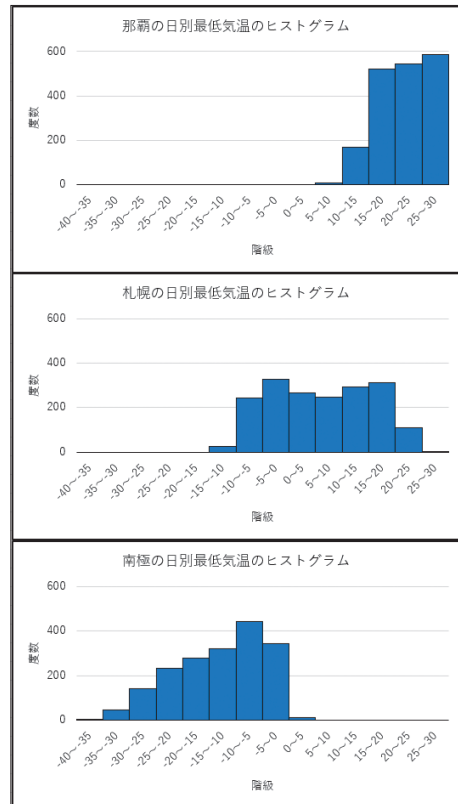
「2-10-4」シートにおいて、図のように5地点の度数分布表を作成しましょう。

絶対参照によるオートフィルを行えば、東京の度数分布表をコピーして、羽田から南極までの度数分布表を作成することができます。挑戦してみましょう。

階級	度数				
	東京	羽田	札幌	那覇	南極
-40~-35	0	0	0	0	6
-35~-30	0	0	0	0	47
-30~-25	0	0	0	0	142
-25~-20	0	0	0	0	232
-20~-15	0	0	0	0	277
-15~-10	0	0	26	0	321
-10~-5	0	0	243	0	444
-5~0	50	17	326	0	345
0~5	389	305	267	0	12
5~10	321	351	247	6	0
10~15	277	283	293	167	0
15~20	342	353	311	521	0
20~25	323	360	111	545	0
25~30	125	158	3	588	0

Step 3 ヒストグラムの作成

Step 2で作成した度数分布表を用いて、図のような那覇・札幌・南極の日別最低気温を比較できるヒストグラムを作成しましょう。



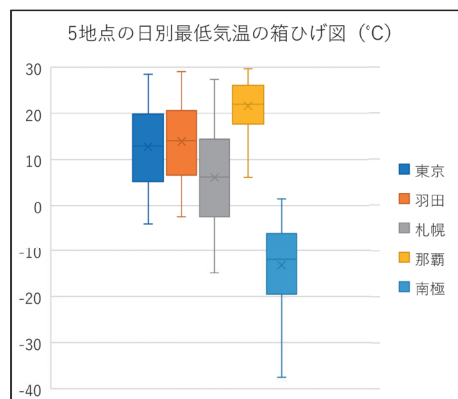
2-10-5 箱ひげ図

Step 1 シートのコピーとシート名の変更

「2-10-1」シートをコピーし、「2-10-5」シートを作成しましょう。

Step 2 箱ひげ図の作成

「2-10-5」シートにおいて、図のように箱ひげ図を作成しましょう。



2-10-6 相関係数と散布図

Step 1 シートのコピーとシート名の変更

「2-10-1」シートをコピーし、「2-10-6」シートを作成しましょう。

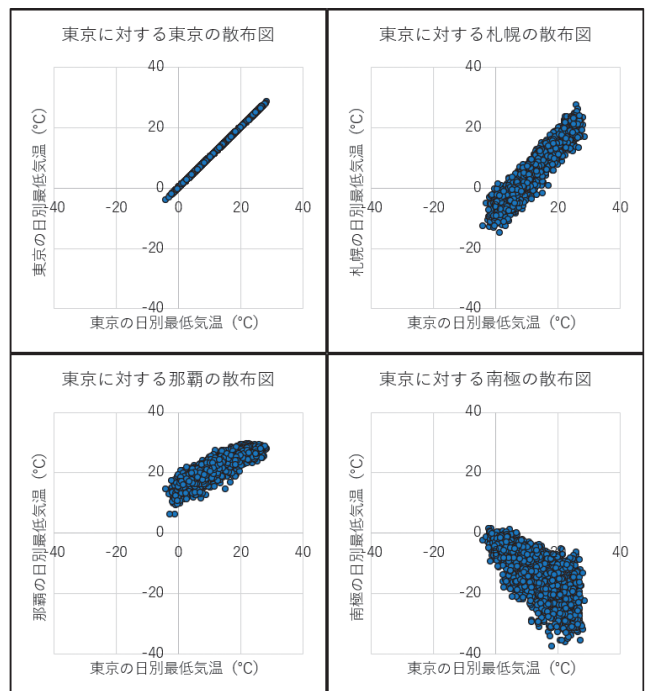
Step 2 度数分布表の作成

「2-10-6」シートにおいて、図のようにB列の東京のデータをC列に複製しましょう。さらに、5地点の最大値・最小値、東京に対する5地点の相関係数を算出しましょう。

	A	B	C	D	E	F	G
1	ダウンロードした時刻：2022/07/22 07:08:41						
2	5地点の日別最低気温（℃）						
3							
4	最大値		28.5	29.0	27.4	29.7	1.4
5	最小値		-4.0	-2.4	-14.9	6.1	-37.5
6	相関係数		1.000	0.996	0.953	0.896	-0.760
7							
8	年月日	東京	東京	羽田	札幌	那覇	南極
9	2016/01/01	3.6	3.6	5.9	-2.6	15.2	-4.4
10	2016/01/02	2.6	2.6	3.9	-0.7	14.8	-4.5
11	2016/01/03	3.0	3.0	4.8	-1.8	19.5	-5.3

Step 3 散布図の作成

東京に対する東京・札幌・那覇・南極の最低気温の散布図を図のように作成しましょう。



2-10-7 データのダウンロードとデータクリーニング2

Step 1 データのダウンロード

気象庁のWebサイト「<https://www.jma.go.jp>」から以下のデータをダウンロードしましょう。

- 地点：札幌
- 項目：天気概況(昼：06時～18時)・日別値・日最低気温
- 期間：2016年1月1日～2020年12月31日

Step 2 ファイルの移動とファイル名の変更

Step 1でダウンロードした「data.csv」ファイルを、「ダウンロード」フォルダーから、2-10-7のStep 1で作成した「2-10」フォルダーに移動しましょう。次に、ファイル名を「札幌の天気概況と最低気温.csv」に変更しましょう。

Step 3 シートの追加とシート名の変更

「札幌の天気概況と最低気温.csv」ファイルの「札幌の天気概況と最低気温」シートを、「2-10.xlsx」ファイルの「2-10-6」シートの右側に追加しましょう。さらに、シートの名前を「札幌の天気概況と最低気温」から、「2-10-7」に変更しましょう。

Step 4 ウィンドウ枠の固定とデータクリーニング

B6でウィンドウ枠を固定し、データクリーニングをして図のようにしましょう。但し、ダウンロードした時刻は揃える必要はありません。

	A	B	C	D
1	ダウンロードした時刻：2022/07/22 08:09:38			
2	札幌の天気概況と最低気温			
3				
4	年月日	天気概況(昼：06時～18時)	最低気温(°C)	
5	2016年01月01日	雪時々曇		-2.6
6	2016年01月02日	みぞれ時々曇一時雨		-0.7
7	2016年01月03日	曇時々晴一時雪		-1.8
8	2016年01月04日	曇一時雪		-3.8
9	2016年01月05日	曇時々雪一時晴		-5.4

2-10-8 クロス集計

Step 1 シートのコピーとシート名の変更

「2-10-7」シートをコピーし、「2-10-8」シートを作成しましょう。

Step 2 クロス集計表の作成

「2-10-8」シートにおいてクロス集計をして、図のように各年にそれぞれの天気概況が何日ずつあったかを示しましょう。

個数 / 最低気温(°C)	列ラベル					総計
	2016年	2017年	2018年	2019年	2020年	
行ラベル						
雨時々曇	1					1
晴時々薄曇		1				1
雪時々曇一時雨			1			1
曇後一時雨		1				1
薄曇一時晴			2			2
みぞれ一時曇後雪		1				1
みぞれ後一時雪					1	1
みぞれ後雨時々曇			1			1
みぞれ後晴一時雨				1		1
みぞれ時々雨後一時晴		1				1
みぞれ時々雪一時曇			1			1
みぞれ時々曇、あられを伴う			1			1
みぞれ時々曇、雷・ひょうを伴う					1	1
みぞれ時々曇一時雨	1					1
みぞれ時々曇一時雪		1				1
計	7	0	4	2	2	25

Step 3 クロス集計の可視化

Step 2で作成したクロス集計表にフィルターをかけて、「雨」・「快晴」・「晴」・「雪」・「曇」の5つの天気概況の日数を可視化したグラフを作成しましょう。

