

目的

- ・平均値, 中央値や最頻値等の基本的なデータの代表値について説明できる。
- ・統計処理用ソフトウェア R を用いて簡単な計算ができる。
- ・統計処理用ソフトウェア R を用いて, ヒストグラム, 箱ひげ図や散布図等を用いて, 簡単なデータの可視化ができる。

使用機器:ノートパソコン

予習事項: 手引書の用語等, 不明なところは予め書籍等を活用して調べておくこと。

準備

(1)Google クラウドにあるファイル「後期 2 年生 R_演習.ipynb」をダウンロードしてください。

(2)下記 URL から今回利用するデータファイルをダウンロードしてください。

気象庁 「過去の気象データ・ダウンロード」 ホームページ

<http://www.data.jma.go.jp/gmd/risk/obsdl/index.php>

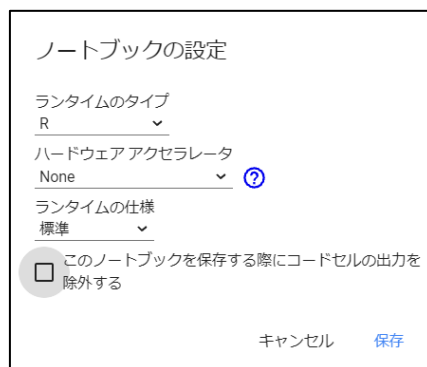
- ① 旭川市の 2011 年 1 月～2020 年 12 月までの月ごとの平均気温(CSV 形式)
- ② 旭川市以外の 2011 年 1 月～2020 年 12 月までの月ごとの平均気温(CSV 形式)

(3) Google 社 Colabrotory <https://colab.research.google.com/> のページを開く

(あ) ページを開いた後, 演習用ファイル「後期 2 年生用 R.ipynb」を手元のパソコンからアップロード
ページ上部から「ファイル」→「ノートブックをアップロード」を選択, ファイルを選べるので演習用のファイルを選び, アップロードしてください。

(い) R の命令が実行可能なことを確認

ページ上部から「ランタイム」→「ノートブックの設定において, 下図のように「ランタイムのタイプ」が「R」に指定されていることを確認してください。できていない方はできている班員の方に相談してみてください。



演習1 「簡単な計算」

別紙の説明用資料にあります四則演算や平方根、対数の計算を試し、同じ結果になることを確認してください。確認できたら、下記の実行結果を得てください。

設問1 2次方程式を解の公式を用いて解いてください。

$$x^2 + 32x + 255 = 0$$

演習2 「平均, 中央値, 四分位点および分散の計算」

別紙の説明用資料にあります平均, 中央値, 四分位点および分散の計算を試し、同じ結果になることを確認してください。確認できたら、下記の実行結果を得てください。

- ・平均：データの総和をデータの個数で割ったもの
- ・中央値：データを昇順（小さいもの→大きいもの）に並べた時に中央にくる値
データの個数が奇数の場合：中央の値
データの個数が偶数の場合：中央の2つの値の平均を取る
- ・最頻値：度数が最も多い階級に対する値
- ・四分位点：中央値でデータを2分割し、その分割した中で値が小さいものと大きいものに分けたもの
 - ・第1四分位点：小さいデータの中での中央値
 - ・第3四分位点：大きいデータの中での中央値
- ・(不偏)分散：平均値からのばらつき具合を計算した数値

設問2 2019年および2018年の旭川市の月ごとの平均気温について、資料と同様に平均, 中央値, 四分位点および分散を求めてください。

演習3 「データの可視化」

別紙の説明用資料にありますグラフを用いたデータの可視化を試し、同じ結果になることを確認してください。確認できたら、下記の実行結果を得てください。

- ・ヒストグラム：度数分布表を横軸に階級区間, 縦軸に度数を取ってグラフにしたもの
- ・階級区間：数値で表現可能なデータを予め決められた単位で区切った区間のこと
- ・度数（あるいは, 頻度）：各階級区間にあるデータの個数
- ・度数分布表：階級区間と度数（頻度）の関係を表にまとめたもの
- ・箱ひげ図：第1四分位点, 中央値, 第3四分位点をテューキーの方式で描画したもの
データのばらつきを比較する時に利用
- ・散布図：2つの値の関係を描画したもの

設問3 2019年および2018年の旭川市の月ごとの平均気温について、ヒストグラム, 箱ひげ図および散布図を求めてください。

演習 4 「データファイルの利用」

別紙の説明用資料にありますプログラムを実行すると外部ファイルの読み込みができます。同じ結果になることを確認してください。確認できたら、下記の実行結果を得てください。

(準備)

気象庁のホームページからダウンロードした気温についてのデータファイルを Google Colab のページにアップロードしてください。

設問 4 予めダウンロードした旭川市以外の都市の月ごとの平均気温について、同様に処理できることを確認してください。

- 演習 4 (設問 4) までを 15 時 30 分までに仕上げてください。
もし途中の場合でも、そこまでの成果をまとめ、ジュピターノートブック形式にて保存してください。(要提出：このファイルが実験結果となります。)
※保存方法：Colaboratory 上部の「ファイル」のメニューから「ダウンロード」→「.ipynb ダウンロード」を選択し、お手元あるいはクラウドのご自分の保存領域に保存してください。

考察

- a. (不偏)分散を求める式を調べ、分散の意味をまとめる。
- b. 箱ひげ図にて出てきた「テューキーの方式」について調べ、その方法をまとめる。
- c. 相関係数を求める式を調べ、相関係数の意味をまとめる。特に、「正の相関」、「負の相関」および「無相関」の違いが分かるように説明してください。

(余裕がある人向け)

- 1. $5x^2 + 2x + 15 = 0$ の場合についても考えてみてください。
 - 2. ある都市の気温と関係が深そうなデータを探し出し、考察 c にある相関係数を計算してみてください。例えば、「消費電力」とか、「飲料水の消費量」など…
- 授業終了時まで、下記の考察内容を調べ、レポートを完成させてください。
 - 提出物は、PDF 形式のレポートと上記の演習結果となります。