# 气候变化数据分析

## 概述

本次实验先对 1951 年至 1980 年这 30 年的气候数据进行了散点图展示和线性回归分析，然后对 1882 年至 2014 年的气候数据进行线性回归分析。

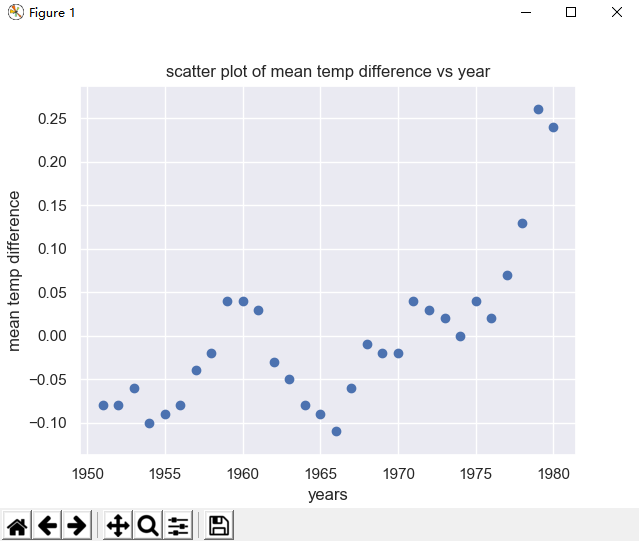
5-year-mean-1951-1980.csv文件包含了1951至1980的气候数据。而 5-year-mean-1882-2014.csv 文件包含了1882年至2014年的气候数据。

## 实验过程

### 创建散点图

加载的数据集使用 1951 年到 1980 年这 30 年的平均值来计算这段期间的基准温度，然后使用 5 年平均温度计算每一年的 5 年平均值和 30 年平均值之差。 散点图显示年温差。

# 创建散点图  
plt.scatter(yearsBase, meanBase)  
plt.title('scatter plot of mean temp difference vs year') # 平均温差与年份的散点图  
plt.xlabel('years', fontsize=12)  
plt.ylabel('mean temp difference', fontsize=12) # 平均温差  
plt.show()

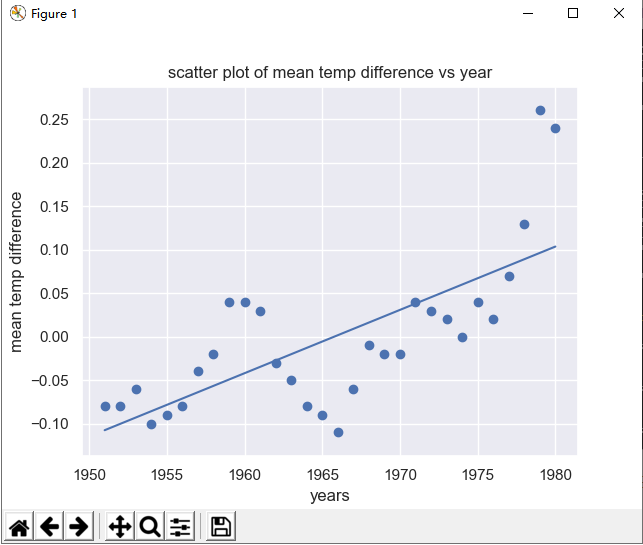


### 使用 Scikit Learn 执行线性回归

[scikit-learn](http://scikit-learn.org/stable/) 擅长构建机器学习模型，以帮助从数据中提取信息。

# 使用 scikit-learn 执行线性回归  
  
# 准备 Linear Regression 模型并实例化它  
model = LinearRegression(fit\_intercept=True)  
  
# 创建模型  
model.fit(yearsBase[:, np.newaxis], meanBase)  
mean\_predicted = model.predict(yearsBase[:, np.newaxis])  
  
# 生成一个图表  
plt.scatter(yearsBase, meanBase)  
plt.plot(yearsBase, mean\_predicted)  
plt.title('scatter plot of mean temp difference vs year') # 平均温差与年份的散点图  
plt.xlabel('years', fontsize=12)  
plt.ylabel('mean temp difference', fontsize=12) # 平均温差  
plt.show()  
  
print(' y = {0} \* x + {1}'.format(model.coef\_[0], model.intercept\_))

显示 sckikit-learn 计算出的带回归线的散点图。

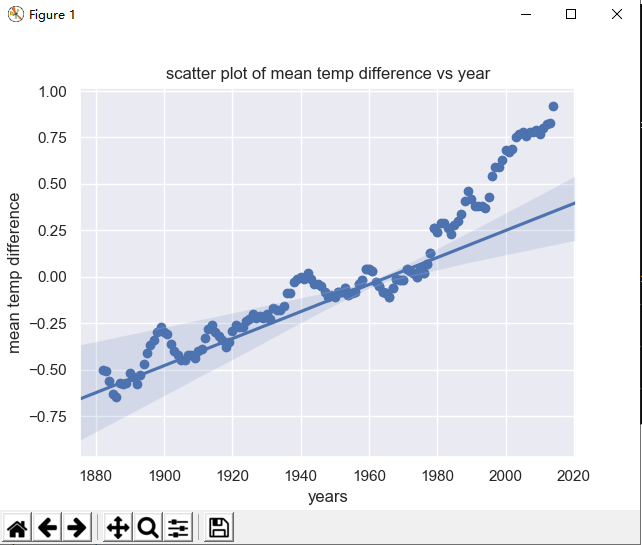


### 使用 Seaborn 分析数据

使用 [Seaborn](https://seaborn.pydata.org/)（用于统计数据可视化的库）为数据集（涵盖自 1882 年至 2014 年的数据）绘图。

Seaborn 可以创建一条伴有投影的回归线，通过一个简单的函数调用根据回归显示数据点应位于什么位置。

# 使用 Seaborn 分析数据  
  
plt.scatter(years, mean)  
plt.title('scatter plot of mean temp difference vs year') # 平均温差与年份的散点图  
plt.xlabel('years', fontsize=12)  
plt.ylabel('mean temp difference', fontsize=12) # 平均温差  
sns.regplot(yearsBase, meanBase)  
plt.show()



## 完整代码

#!/usr/bin/env python  
# coding: utf-8  
  
# 气候数据分析  
  
import matplotlib.pyplot as plt  
import numpy as np  
from sklearn.linear\_model import LinearRegression  
import seaborn as sns; sns.set()  
  
# 导入数据  
  
yearsBase, meanBase = np.loadtxt('./resources/5-year-mean-1951-1980.csv', delimiter=',', usecols=(0, 1), unpack=True)  
years, mean = np.loadtxt('./resources/5-year-mean-1882-2014.csv', delimiter=',', usecols=(0, 1), unpack=True)  
  
# 创建散点图  
  
plt.scatter(yearsBase, meanBase)  
plt.title('scatter plot of mean temp difference vs year') # 平均温差与年份的散点图  
plt.xlabel('years', fontsize=12)  
plt.ylabel('mean temp difference', fontsize=12) # 平均温差  
plt.show()  
  
# 使用 scikit-learn 执行线性回归  
  
# 准备 Linear Regression 模型并实例化  
model = LinearRegression(fit\_intercept=True)  
  
# 创建模型  
model.fit(yearsBase[:, np.newaxis], meanBase)  
mean\_predicted = model.predict(yearsBase[:, np.newaxis])  
  
# 生成一个图表  
plt.scatter(yearsBase, meanBase)  
plt.plot(yearsBase, mean\_predicted)  
plt.title('scatter plot of mean temp difference vs year') # 平均温差与年份的散点图  
plt.xlabel('years', fontsize=12)  
plt.ylabel('mean temp difference', fontsize=12) # 平均温差  
plt.show()  
  
print(' y = {0} \* x + {1}'.format(model.coef\_[0], model.intercept\_))  
  
  
# 使用 Seaborn 分析数据  
  
plt.scatter(years, mean)  
plt.title('scatter plot of mean temp difference vs year') # 平均温差与年份的散点图  
plt.xlabel('years', fontsize=12)  
plt.ylabel('mean temp difference', fontsize=12) # 平均温差  
sns.regplot(yearsBase, meanBase)  
plt.show()

## 数据来源

5-year-mean-1951-1980.csv 和 5-year-mean-1882-2014.csv 文件来源于 [此链接](https://a4r.blob.core.windows.net/public/notebook-resources.zip) ，由微软提供。