МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«Национальный исследовательский ядерный университет «МИФИ» (НИЯУ МИФИ)

Институт Финансовых Технологий и Экономической Безопасности Кафедра Финансового мониторинга

**Лабораторная работа №1 по курсу**

**«Макростатический анализ и прогнозирование»**

|  |  |
| --- | --- |
| **Выполнил студент группы С21-703:** | Монастырский М.О. |
| **Проверил:** | Домашова Д.В. |

Москва, 2024

Оглавление

[1. Постановка задачи 3](#_Toc178717567)

[2. Выполнение работы 4](#_Toc178717568)

[2.1 Метод полных связей 4](#_Toc178717569)

[2.2 Метод Уорда 10](#_Toc178717570)

[2.3 Метод одиночной связи 13](#_Toc178717571)

[2.4 Метод k-средних 17](#_Toc178717572)

1. Постановка задачи

1. Выбор предмета исследования. В данной лабораторной работе предметом исследования является состояние субъектов РФ, которое характеризуется следующими показателями:

|  |  |
| --- | --- |
|  | Число дорожно-транспортных происшествий и пострадавших в них на 100 000 человек населения |
|  | Смертность населения старше трудоспособного возраста, на 100 000 человек населения соответствующего возраста |
|  | Продажа сильно алкогольной продукции населению(тысяч декалитров)/**на тыс населения** |
|  | Средняя Стоимость минимального (условного) набора потребительских товаров и услуг |
|  | Число спортивных сооружений/ **на тыс населения** |
|  | Доходы консолидированных бюджетов субъектов Российской Федерации / **на тыс населения** |
|  | Предварительно расследовано преступлений, совершенных в состоянии алкогольного опьянения/ **на тыс населения** |
|  | Среднедушевые доходы населения (в месяц), руб. |
|  | Численность студентов, обучающихся по программам бакалавриата, специалитета, магистратуры на 10 000 человек населения, всего |

1. По данным Приложения А c помощью методов кластерного анализа:

* провести классификацию субъектов РФ с помощью иерархических агломеративных методов кластерного анализа;
* провести классификацию субъектов РФ с помощью метода К-средних.

1. Сравнить классификации, полученные с помощью агломеративных кластер-процедур и метода К-средних, обосновать выбор окончательного варианта классификации;
2. Дать экономическую интерпретацию результатов классификации

2. Выполнение работы

2.1 Метод полных связей

Методом «полных связей» при пороговом значении расстояния все субъекты РФ разбиваются на 7 классов S1= {S11, S12, S13, S14, S15, S16,S17}, состав которых приведен в таблице 1.

|  |  |  |
| --- | --- | --- |
| **Номер кластера** | **Количество объектов** | **Состав класса** |
| {S11} | 2 | Ненецкий Автономный Округ, Чукотский Автономный Округ |
| {S12} | 3 | Томская область, Санкт-Петербург, Москва |
| {S13} | 4 | Кабардино-Балкарская Республика  Республика Дагестан  Республика Ингушетия  Чеченская Республика |
| {S14} | 26 | Алтайский Край, Амурская Область, Архангельская Область, Брянская Область, Владимирская Область, Вологодская Область, Еврейская Автономная Область, Иркутская Область, Калининградская Область, Кировская Область, Костромская Область, Нижегородская Область, Оренбургская Область, Пермский Край, Приморский Край, Псковская Область, Республика Бурятия, Республика Карелия, Республика Коми, Республика Крым, Республика Саха, Республика Хакассия, Сахалинская Область, Удмуртская Республика, Хабаровский Край, Ярославская Область |
| {S15} | 4 | Камчатский Край, Магаданская Область, Ханты-Мансийский Автономный Округ, Ямало-ненецкий Автономный Округ |
| {S16} | 29 | Астраханская Область, белгородская Область, Волгоградская Область, Воронежская Область, Ивановская Область, Калужская Область, Карачаево-Черкесская Республика, Краснодарский Край, Красноярский Край, Ленинградская область, Московская Область, Мурманская Область, Новосибирская Область, Орловская Область, Республика Адыгея, Республика Марий-Эл, Республика Мордовия, Республика Северная Осетия – Алания, Республика Татарстан, Ростовская Область, Рязанская Область, Самарская Область, Саратовская Область, Свердловская Область, Севастополь, Ставропольский Край, Тульская Область, Ульяновская Область, Челябинская Область |
| {S17} | 17 | Забайкальский Край, Кемеровская Область, Курганская Область, Курская Область, Липецкая Область, Новгородская Область, Омская Область, Пензенская Область, Республика Алтай, Республика Башкортостан, Республика Калмыкия, Республика Тыва, Смоленская Область, Тамбовская область, Тверская Область, Тюменская Область, Чувашская Республика |



**Экономическая интерпретация**

Первый кластер характеризуется наиболее низким числом ДТП и наиболее низким числом студентов ВУЗов. В противовес этому первый кластер выделяется самой высокой смертностью населения старше трудоспособного возраста, самой высокой продажей крепкого алкоголя, числом спортивных сооружений на душу населения, самыми высокими доходами субъекта, самым большим количеством предварительно расследованных преступлений, совершенных в состоянии алкогольного опьянения, и самыми высокими среднедушевыми доходами

Второй кластер выделяется самым высоким количеством студентов

Третий кластер выделяется самым низким числом ДТП, самой низкой смертностью населения старше трудоспособного возраста, самым трезвым образом жизни, самой дорогой потребительской корзиной, самым низким числом спортивных сооружений, самыми маленькими доходами бюджетов, самым низким количеством расследованных преступлений, совершенных в состоянии алкогольного опьянения и самыми низкими среднедушевыми доходами. Четвертый, Шестой и Седьмой кластеры имеют средние показатели во всех областях. Пятый кластер представляет собой регионы с минимальной стоимостью товаров и услуг

Кластер 1

Этот кластер, включает северные или восточные регионы с суровыми климатическими условиями. Высокие доходы могут быть связаны с добычей полезных ископаемых или других природных ресурсов. Однако трудные климатические условия и изолированность могут ограничивать образовательные возможности и способствовать высокому потреблению алкоголя, что ухудшает показатели по смертности.

Кластер 2

Кластер с высоким количеством студентов включает региональные центры и города с развитой образовательной инфраструктурой, такие как Московская и Санкт-Петербургская области. Географически удобное положение и высокая концентрация вузов способствуют привлечению студентов. Это, в свою очередь, стимулирует экономику и снижает уровень преступности.

Кластер 3

Третий кластер может охватывать южные регионы с развитым сельским хозяйством и благоприятным климатом, способствующим здоровому образу жизни. Низкие показатели по алкоголю и ДТП могут быть связаны с высокой культурой ведения здорового образа жизни. Однако малый доход бюджета и высокая стоимость жизни могут указывать на отсутствие инвестиций в экономику, несмотря на хорошие социальные показатели.

Кластеры 4, 6 и 7

Эти кластеры, состоят из регионов, которые находятся в средней полосе и обладают сбалансированной структурой экономики. Безопасность и стабильность характерны для этих территорий, где нет доминирующих экстремальных показателей. Это может говорить о равномерном развитии инфраструктуры и ресурсов.

Кластер 5

Пятый кластер с минимальными ценами на товары и услуги, возможно, включает регионы с сельскохозяйственной ориентацией или периферийные территории, где низкая стоимость жизни компенсируется отсутствием крупных индустрий и капиталовложений. Географическая изоляция или удаленность от крупных экономических центров может способствовать снижению стоимости жизни, но также ограничивать экономический рост.

Кластер 1 **Кластер "Алкогольная зависимость и высокая смертность"**

**Кластер 2 Студенческий**

**Кластер 3 Бедные ЗОЖники**

**Кластер 4,6,7 Средние показатели**

**Кластер 5 "Доступные товары и услуги"**

2.2 Метод Уорда

Методом «Уорда» при пороговом значении расстояния все субъекты РФ разбиваются на 5 классов S2= {S21, S22, S23, S24,S25}, состав которых приведен в таблице 2.

|  |  |  |
| --- | --- | --- |
| **Номер кластера** | **Количество объектов** | **Состав класса** |
| {S21} | 4 | Кабардино-Балкарская Республика, Республика Дагестан, Республика Ингушетия, Чеченская Республика |
| {S22} | 42 | Архангельская Область, Астраханская Область, Брянская Область, Владимирская Область, Волгоградская Область, Вологодская Область, Ивановская Область, Иркутская Область, Калининградская Область, Калужская Область, Кировская Область, Костромская Область, Краснодарский Край, Красноярский Край, Ленинградская Область, Московская Область, Мурманская Область, Нижегородская Область, Новосибирская Область, Оренбургская Область, Орловская Область, Пермский Край, Приморский Край, Псковская Область, Республика Бурятия, Республика Крым, Республика Марий-Эл, Республика Мордовия, Республика Саха, Республика Татарстан, Республика Хакассия, Рязанская Область, Самарская Область, Саратовская область, Свердловская Область, Севастополь, Тульская Область, Удмуртская Республика, Ульяновская Область, Хабаровский Край, Челябинская Область, Ярославская Область |
| {S23} | 6 | Камчатский Край, Магаданская Область, Ненецкий Автономный Округ, Ханты-Мансийский Автономный Округ, Чукотский Автономный Округ, Ямало-ненецкий Автономный Округ, |
| {S24} | 14 | Белгородская Область, Воронежская Область, Москва, Карачаево-Черкесская Область, Курская Область, Омская Область, Республика Адыгея, Республика Калмыкия, Республика Северная Осетия – Алания, Ростовская Область, Санкт-Петербург, Ставропольский Край, Томская Область, Тюменская Область. |
| {S25} | 19 | Алтайский Край, Амурская Область, Еврейская Автономная Область, Забайкальский Край, Кемеровская Область, Курганская Область, Липецкая Область, Новгородская Область, Пензенская Область, Республика Алтай, Республика Башкортостан, Республика Карелия, Республика Коми, Республика Тыва, Сахалинская Область, Смоленская Область, Тамбовская Область, Тверская Область, Чувашская Республика |

з

**Экономическая интерпретация**

Первый кластер характеризуется самыми низкими показателями по всем характеристикам, за исключением самых высоких цен на минимальный набор товаров и услуг.

Второй Кластер Имеет средние показатели по всем параметрам

Третий Кластер выделяется самыми высокими показателями продажи алкоголя, самыми низкими ценами на товары и услуги, Самыми высокими доходами населения и субъекта, самым высоким количеством предварительно расследованных преступлений и самым низким числом студентов.

Четвертый кластер характеризуется самым высоким числом студентов

Пятый кластер характеризуется самым большим числом ДТП, самой высокой смертностью людей старше трудоспособного возраста и самым высоким числом спортивных сооружений на душу.

1 Кластер **"Низкие показатели и высокие цены"**

**2 Кластер «Средние показатели»**

**3 Кластер «Алкогольный рай с низкими ценами»**

**4 Кластер Студенческий**

**5 Кластер Высокая смертность**

2.3 Метод одиночной связи

Методом «одиночной связи» при пороговом значении расстояния все субъекты РФ разбиваются на 7 классов S1= {S31, S32, S33, S34, S35, S36;S37}, состав которых приведен в таблице 3.

|  |  |  |
| --- | --- | --- |
| **Номер кластера** | **Количество объектов** | **Состав класса** |
| {S31} | 1 | Ненецкий Автономный Округ |
| {S32} | 1 | Республика Ингушетия |
| {S33} | 1 | Ханты-Мансийский Автономный округ |
| {S34} | 1 | Чукотский Автономный Округ |
| {S35} | 1 | Ямало-Ненецкий Автономный Округ |
| {S36} | 2 | Санкт-Петербург, Москва |
| {S37} | 78 | Алтайский Край, Амурская Область, Архангельская Область, Астраханская Область, Белгородская Область, Брянская Область, Владимирская Область, Волгоградская Область, Вологодская Область, Воронежская Область, Еврейская Автономная Область, Забайкальский Край, Ивановская Область, Иркутская Область, Кабардино-Балкарская Республика, Калининградская область, Калужская область, Камчатский Край, Карачаево-Черкесская республика, Кемеровская Область, Кировская Область, Костромская Область, Краснодарский Край, Красноярский Край, Курганская Область, Курская Область, Ленинградская Область, Липецкая Область, Магаданская Область, Московская Область, Мурманская Область, Нижегородская Область, Новгородская Область, Новосибирская Область, Омская Область, Оренбургская Область, Орловская Область, Пензенская Область, Пермский Край, Приморский Край, Псковская Область, Республика Адыгея, Республика Алтай, Республика Башкортостан, Республика Бурятия, Республика Дагестан, Республика Калмыкия, Республика Карелия, Республика Коми, Республика Крым, Республика Марий-Эл, Республика Мордовия, Республика Саха, Республика Северная Осетия – Алания, Республика Татарстан, Республика Тыва, Республика Хакассия, Ростовская Область, Рязанская Область, Самарская Область, Саратовская Область, Сахалинская Область, Свердловская Область, Севастополь, Смоленская Область, Ставропольский Край, Тамбовская Область, Тверская Область, Томская Область, Тульская Область, Тюменская Область, Удмуртская Республика, Ульяновская Область, Хабаровский Край, Челябинская Область, Чеченская Республика, Чувашская Республика, Ярославская Область. |



**Экономическая Интерпретация**

Кластер номер 1 характеризуется самым большим числом спортивных сооружений и самыми высокими доходами субъектов.

Кластер номер 2 характеризуется самой низкой смертностью населения старше работоспособного возраста, Самыми низкими продажами алкоголя, Самой высокой стоимостью минимальной потребительской корзины, Самым низким количеством спортивных сооружений, Самыми низкими доходами субъектов, самыми низкими предварительными расследованиями.

Кластер номер 3 выделяется самой дешевой потребительской корзиной

Кластер 4 выделяется Самым низким числом ДТП, Самым высоким числом смертности, самой большой продажей алкоголя.

Кластер 5 характеризуется самым маленьким числом студентов

Кластер 6 характеризуется самым большим числом студентов

Кластер 7 характеризуется наиболее усредненными показателями

Кластер 1 **«Спортивная утопия»**

Кластер 2 **«Денег нет но вы держитесь»**

Кластер 3 **«Доступные товары**

Кластер 4 **«Алкогольные смерти»**

Кластер 5 **«Дураки»**

Кластер 6 **«Ученые»**

Кластер 7 **«Средние показатели»**

2.4 Метод k-средних

Использование различных методов иерархического агломеративного кластерного анализа приводит к различным результатам классификации. Метод k-средних позволяет получить более устойчивое разбиение, но требует задания некоторых начальных условий (количество образуемых кластеров, порог завершения процесса классификации и т. д.).

|  |  |  |
| --- | --- | --- |
| **Номер кластера** | **Количество объектов** | **Состав класса** |
| {S41} | 32 | Астраханская Область, Брянская Область, Волгоградская Область, Воронежская Область, Ивановская Область, Иркутская Область, Калининградская Область, Калужская Область, Краснодарский Край, Красноярский Край, Курская Область, Нижегородская Область, Новосибирская Область, Омская Область, Оренбургская Область, Орловская Область, Пермский Край, Республика Крым, Республика Марий-Эл, Республика Мордовия, Республика Саха, Республика Татарстан, Рязанская Область, Самарская Область, Саратовская Область, Свердловская Область, Томская Область, Тульская Область, Удмуртская Республика, Ульяновская Область, Челябинская Область, Ярославская Область. |
| {S42} | 3 | Ненецкий Автономный Округ, Чукотский Автономный Округ, Ямало-ненецкий Автономный округ |
| {S43} | 17 | Алтайский Край, Белгородская Область, Кемеровская Область, Курганская Область, Липецкая Область, Новгородская Область, Пензенская Область, Республика Башкортостан, Республика Бурятия, Республика Калмыкия, Республика Тыва, Республика Хакассия, Смоленская Область, Тамбовская Область, Тверская Область, Тюменская Область, Чувашская республика |
| {S44} | 14 | Москва, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Московская область, Республика Адыгея, Республика Дагестан, Республика Ингушетия, Республика Северная Осетия – Алания, Ростовская Область, Санкт-Петербург, Севастополь, Ставропольский край, Ханты-Мансийский Автономный округ, Чеченская Республика |
| {S25} | 19 | Амурская Область, Архангельская Область, Владимирская Область, Вологодская Область, Еврейская Автономная Область, Забайкальский Край, Камчатский Край, Кировская Область, Костромская Область, Ленинградская Область, Магаданская Область, Мурманская Область, Приморский Край, Псковская Область, Республика Алтай, Республика Карелия, Республика Коми, Сахалинская Область, Хабаровский Край. |

**Экономическая интерпретация**

Кластер 1 характеризуется самым большим числом студентов

Кластер 2 характеризуется Самым маленьким числом ДТП, самой маленькой стоимостью набора услуг,самыми высокими доходами субъекта, самыми высокими показателями предварительной раскрываемости, самыми высокими среднедушевыми доходами, при этом самыми низкими показателями количества студентов

Третий кластер характеризуется самыми низкими среднедушевыми доходами населения

Четвертый кластер выделяется самыми низкими показателями ДТП, самыми низкими показателями смертности, Самой низкой продажей алкоголя, Самым низким количеством спортивных сооружений, самым низким количеством предварительно расследованных преступлений.

Пятый кластер выделяется самым большим количеством ДТП, самой большой смертностью населения старше трудоспособного возраста, самыми дорогими услугами и товарами, и самым маленьким числом студентов.

Кластер 1 **«Студенческий»**

Кластер 2 **«Высокие доходы и безопасные дороги»**

Кластер 3 **«Низкие доходы»**

Кластер 4 **«Показательное дно»**

Кластер 5 **«Проблемы с безопасностью образованием и ценами»**

#!/usr/bin/env python

# coding: utf-8

# In[4]:

get\_ipython().system('pip install openpyxl')

# In[163]:

import pandas as pd

from sklearn.preprocessing import StandardScaler, MinMaxScaler

from scipy.cluster.hierarchy import linkage, dendrogram,fcluster

from sklearn.cluster import KMeans

from matplotlib import pyplot as plt

# In[44]:

# Load data

df = pd.read\_excel("data.xlsx",sheet\_name="2022",index\_col = "Наименование")

# In[45]:

df.drop("Unnamed: 0",axis=1,inplace=True)

# In[46]:

df["Доходы консолидированных бюджетов субъектов Российской Федерации (млн)"].replace(

                                regex={',': '.'," ":"",'–':0},inplace=True)

# In[47]:

df\_filtered = pd.DataFrame({"X1":df["Число дорожно-транспортных происшествий и пострадавших в них на 100 000 человек населения"],

                            "X2":df["Смертность населения старше трудоспособного возраста, на 100 000 человек населения соответствующего возраста"],

                           "X3":df["Продажа сильно алкогольной продукции населению(тысяч декалитров)"].div(df["Численность населения, тыс. человек"]),

                            "X4":df["Стоимость минимального (условного) набора потребительских товаров и услуг"].astype(float),

                            "X5":df["Число спортивных сооружений"].div(df["Численность населения, тыс. человек"]),

                            "X6":df["Доходы консолидированных бюджетов субъектов Российской Федерации (млн)"].astype(float).div(df["Численность населения, тыс. человек"]),

                            "X7":df["Предварительно расследовано преступлений, совершенных в состоянии алкогольного опьянения"].div(df["Численность населения, тыс. человек"]),

                            "X8":df["Среднедушевые доходы населения (в месяц), руб."],

                            "X9":df["Численность студентов, обучающихся по программам бакалавриата, специалитета, магистратуры на 10 000 человек населения, всего"].astype(float)

                           })

# In[48]:

df\_filtered

# In[ ]:

# In[ ]:

# In[49]:

df\_filtered.dtypes

# In[50]:

df\_filtered["X6"] - df\_filtered["X6"].mean(axis=0)/df\_filtered["X6"]

# In[51]:

for col in df\_filtered.columns:

    sdf[col] = (df\_filtered[col] - df\_filtered[col].mean(axis=0))/df\_filtered[col].std()

# In[52]:

sdf["Наименование"] = df\_filtered.index

# In[53]:

sdf.set\_index("Наименование",inplace=True)

# In[54]:

sdf

# In[56]:

sdf["X6"] = [-0.832256599,

0.517564476,

0.155036569,

-0.359749214,

-0.222913948,

-0.403474737,

-0.406735079,

-0.492329036,

-0.282278512,

-0.386556673,

0.73188267,

0.434812446,

1.34589173,

-0.388519731,

-0.653255706,

-0.543854591,

-0.291236648,

-0.267983437,

2.0575036,

-0.322220544,

0.041703624,

-0.321621748,

-0.316242875,

-0.420316319,

-0.525390417,

-0.326985368,

-0.324499661,

-0.277539357,

-0.373201267,

2.0997129,

-0.188240646,

0.293967434,

4.95772489,

-0.309071346,

-0.260027907,

-0.339015896,

-0.0558568362,

-0.34632205,

-0.326353313,

-0.394927281,

-0.270026224,

-0.191349951,

-0.280892029,

-0.490077504,

0.140997861,

-0.419513275,

0.232711785,

-0.583393156,

-0.392435482,

-0.408194646,

-0.117387142,

0.298226211,

-0.990065046,

-0.430188766,

-0.27650776,

-0.17593093,

-0.450680697,

-0.202190013,

1.34630413,

-0.393110994,

-0.427326287,

-0.34977747,

-0.249657056,

0.131735835,

-0.491780111,

-0.990065046,

-0.2559698,

-0.990065046,

-0.356003993,

-0.509813052,

-0.305180798,

-0.275965286,

0.87357587,

-0.345618213,

0.211570693,

-0.351707674,

-0.460731175,

0.263057386,

0.713163046,

-0.40467678,

-0.337396802,

-0.43700844,

4.80229078,

3.11894892,

-0.198719471]

# In[58]:

sdf

# In[208]:

def occurrences(s, lst):

    #print(s,lst)

    r = []

    for i in range (0, len(lst)):

        if(lst[i]==s):

            r.append(i)

    return r

def get\_obj\_names(clustering:list, labels:list) -> list:

    Max\_classes = max(clustering)

    res = [[] for i in range(Max\_classes+1)]

    for i in range(0,Max\_classes+1):

        t = occurrences(i,clustering)

        #print(t)

        for reg in t:

            res[i].append(labels[reg])

    return res

method = ["single","complete","ward"]

treshhold = [7,7,5]

r = dict()

for m in range (0,len(method)):

    Z = linkage(sdf,method[m])

    fig = plt.figure(figsize=(25, 10))

    dn = dendrogram(Z,labels=sdf.index)

    plt.show()

    clusters = fcluster(Z,treshhold[m], criterion='maxclust')

    print(f"{method[m]}")

    r[method[m]]=get\_obj\_names(clusters,sdf.index.tolist())[1:]

    for cl in get\_obj\_names(clusters,sdf.index.tolist())[1:]:

        print(f"{cl}", end="\n")

# In[209]:

r

# In[210]:

r["single"]

# In[211]:

means = {"ward":list(),"complete":list(),"single":list(),"kmeans":list()}

for m in method:

    for cl in r[m]:

        #print(cl)

        data =sdf[sdf.index.isin(cl)].mean().tolist()

        means[m].append(data)

        #print(means)

i =0

for c in means["ward"]:

    i+=1

    plt.scatter([range(1,10)],c,label=str(i))

    plt.legend()

    #print(c)

# In[212]:

i =0

for c in means["complete"]:

    i+=1

    plt.scatter([range(1,10)],c,label=str(i))

    plt.legend()

    #print(c)

# In[213]:

i =0

for c in means["single"]:

    i+=1

    plt.scatter([range(1,10)],c,label=str(i))

    plt.legend()

    #print(c)

# In[214]:

def QF(center, data):

    s = 0

    for point in data:

        s+= sum((a-b)\*\*2 for a,b in zip(center,point))

    return s

# In[215]:

for m in method:

    s= 0

    for cl in r[m]:

        s+= QF(sdf[sdf.index.isin(cl)].mean().tolist(),sdf[sdf.index.isin(cl)].values.tolist())

    print(m,s)

# In[217]:

kmeans = KMeans(n\_clusters = 5, random\_state=17)

kmeans.fit(sdf)

# In[218]:

km\_labels = get\_obj\_names(kmeans.labels\_,sdf.index.tolist())[1:]

# In[219]:

for cl in km\_labels:

        #print(cl)

        data =sdf[sdf.index.isin(cl)].mean().tolist()

        means["kmeans"].append(data)

# In[220]:

i =0

for c in means["kmeans"]:

    i+=1

    plt.scatter([range(1,10)],c,label=str(i))

    plt.legend()

    #print(c)

# In[228]:

s=0

for cl in km\_labels:

    s+= QF(sdf[sdf.index.isin(cl)].mean().tolist(),sdf[sdf.index.isin(cl)].values.tolist())

print("kmeans",s)

# In[ ]:

# In[ ]:

# In[ ]: