Виды анализа и R Markdown

Никитин Никита

25 11 2021

## Разведочный анализ данных

### Пример

Согласно [определению](http://domic.isu.ru/student/entity/5345/content/analysis.pdf), цель разведочного анализа данных – изучить данные и найти взаимосвязи, о которых ранее не было известно. Разведочный анализ данных:

* Изучает как могут быть связаны различные переменные.
* Полезно для обнаружения новых связей.
* Помогает формулировать гипотезы и управлять планированием будущих исследований и сбора данных.

[Также](https://ru.wikipedia.org/wiki/%D0%A0%D0%B0%D0%B7%D0%B2%D0%B5%D0%B4%D0%BE%D1%87%D0%BD%D1%8B%D0%B9_%D0%B0%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85) одной из целей разведочного анализа данных является **обнаружение отклонений и аномалий** в данных.

Рассмотрим пример [исследования](<https://cyberleninka.ru/article/n/bezrabotitsa-kak-odna-iz-globalnyh-problem-sovremennogo-mira>, В статье проведен анализ уровня безработицы за 2015-2017 гг., найдены различные связи различных переменных, обозреваются различные причины безработицы:

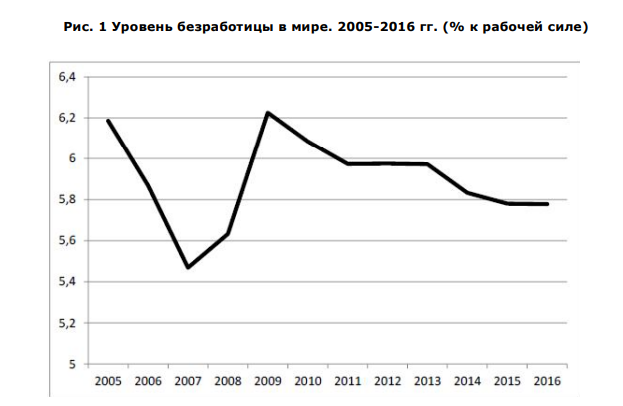
— Уровень безработицы – экономический показатель, один из основных индикаторов на рынке труда. Он отражает соотношение между нетрудоустроенными гражданами в обществе и теми, кто имеет постоянную работу.

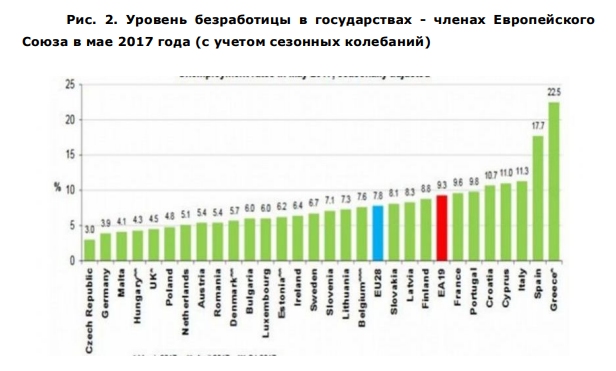
Иными словами, требовалось выявить возможные зависимости безработицы в мире от определённых факторов.

В статье используются данные Росстата, Евростата и ОЭСР, проблема безработицы анализируется на международном уровне.

Безработица, незащищенная занятость и работающие бедные – тенденции и прогнозы на 2016-18 гг.:

|  |  |  |  |
| --- | --- | --- | --- |
| млн. чел / год | 2016 | 2017 | 2018 |
| В мире | 197,7 | 201,1 | 203,8 |
| Развитые страны | 38,6 | 37,9 | 38 |
| Страны с формирующимся рынком | 143,4 | 147 | 149 |
| Развивающиеся страны | 15,7 | 16,1 | 16,6 |

В рамках исследования была построена линейный график уровня безработицы: 



В ходе исследования рассматривались страны:

1. Англия;
2. Германия;
3. Швецария;
4. Ирландия;
5. Испания;
6. Греция;

### Выводы

Исследование использует методы **разведочного анализа данных**, чтобы обнаружить **взаимосвязи между переменными, о которых заранее неизвестно** (например, взаимосвязь между переменной, характеризующей год, и переменной, характеризующей процент безработицы).

## Механистический анализ данных

### Пример

Согласно [определению](http://domic.isu.ru/student/entity/5345/content/analysis.pdf), цель механистического анализа данных – понять, какие именно изменения в переменных приводят к точным изменениям в других переменных. Механистический анализ данных применяется для:

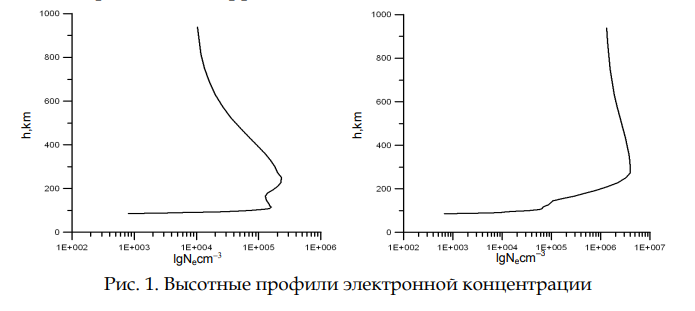
* Применяется к простым ситуациям или ситуациям, которые хорошо моделируются детерминированными уравнениями.
* Обычно применяется к физическим или инженерным наукам.
* Например: биологические науки слишком «шумны» для использования механистического анализа.
* Часто единственный шум в данных – ошибки измерения.

Рассмотрим пример [исследования](https://cyberleninka.ru/article/n/chislennaya-model-rascheta-radiotrass-korotkih-radiovoln-v-ionosfere), целью которого являлся проведение расчетов радиотрасс в авроральной ионосфере и поглощения радиоволн вдоль них в зависимости от выбора геофизических условий.

Использовано уравнение баланса процессов ионизации и рекомбинации в ионосфере:

где α — коэффициент рекомбинации; Ne0 — фоновая концентрация электронов по модели IRI; ∆q — функция корпускулярной ионизации молекул нейтральной атмосферы авроральными электронами, высыпающимися из магнитосферы в ионосферу; q0 — фоновое значение функции ионизации.

Некоторые из результатов расчетов получены для зимних условий 22.12.1969 г. при среднем уровне солнечной и геомагнитной активности (F10,7 = 150, Ap = 27). Для этих условий обнаружено более слабое развитие дневного слоя ионосферы по модели IRI по сравнению с результатами модельных расчетов и данными экспериментальных ионо- грамм для радиотрассы Мурманск — Санкт-Петербург . На основе данных по параметрам и из [9] проведена дополнительная коррекция высотных профилей электронной концентрации модели . На рисунке слева показан пример профиля Ne(h) модели IRI над станцией (65°N, 290°E) в указанных условиях для мирового времени 16,6 ч, а справа — его коррекция



### Выводы

Поскольку рассмотренное исследование имело дело с ситуацией, которая может быть смоделирована с помощью уравнения, а также обнаружило определённую связь между переменными с высоким уровнем доверия, можно считать, что исследование использовано механистический анализ данных.