# 

# 

# 

# Описание данных

Анализировались данные о заработной плате и образовании молодых мужчин в США с 1980 по 1987гг.

Данные содержат 4360 записей и 12 переменных.

# Описательные статистики

Характеристики метрических переменных исследуемого набора данных представлены в таблице

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Переменная | Минимум | Среднее | Медиана | Максимум |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

# T-тест

# Тест Манна-Уитни

Переменная <x> принимает всего <k> значений, поэтому для выявления различий по <x> по показателю <y> используем критерий Манна-Уитни. Согласно этому критерию <не> выявлены статистически значимые различия между <группа1> и <группа2> (p=<p-value>, W = <значение критерия>, медианные значения <медиана в группе 1> и <медиана в группе 2> соответственно).

# Хи-квадрат

Для выявления взаимосвязей между номинальными переменными используется критерий хи-квадрат. Так, в рассматриваемых данных показано <наличие/отсутствие> статистически значимой взаимосвязи между <x> и <y> (p=<p-value>, X = <chi>).

# ANOVA

Рассмотрим различия в уровне <x> по показателю <y>, разделяющему выборку на <k> групп. Согласно анализу ANOVA статистически значимые различия между группами <не> выявлены (p=<p-value>, F = <значение критерия>).

# Корреляционный анализ

Корреляционный анализ позволяет определить взаимосвязь между метрическими переменными. Значения коэффициентов корреляции представлены в таблице, статистически значимые взаимосвязи выделены полужирным шрифтом

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | <var1> | <var2> | <var3> | <var4> |
| <var1> | - | <> | <> | <> |
| <var2> | <> | - | <> | <> |
| <var3> | <> | <> | - | <> |
| <var4> | <> | <> | <> | - |

В частности, <какая-то переменная> <прямо/обратно> связана с <такой-то переменной>, причем связь <достаточно сильная/очень сильная/средней силы> (r = <коэффициент корреляции>, p = <p-value>).

# Корреляционная плеяда

Результаты корреляционного анализа можно визуализировать в виде корреляционной плеяды, представленной ниже.

<рисунок>

# Регрессионный анализ

Регрессионный анализ позволяет определить зависимость между <y> и такими переменными, как <x1>, <x2>, <x3>. Значения коэффициентов регрессионного уравнения и уровни значимости представлены в таблице

|  |  |  |  |
| --- | --- | --- | --- |
|  | коэффициент | t-критерий | p-value |
| <var1> | <> | <> | <> |
| <var2> | <> | <> | <> |
| <var3> | <> | <> | <> |
| <var4> | <> | <> | <> |

Уравнение регрессии выглядит следующим образом:

<уравнение>

То есть изменение <x1> не одну единицу приведет к изменению <y> на <k> единиц.

Отметим, что данная модель объясняет <r>% изменчивости данных.

# Кластеризация

<придумать свое описание>