

# HW - Hypothesis

## Задача

Выдвинуть (или взять уже готовую) ~~стереотип~~ гипотезу и доказать ее с точки зрения статистики

## Шаги

### Выбор гипотезы

Мы на ПМИ часто сталкиваемся с неравномерным распределением девочек и мальчиков, в то время, как на просторах кампуса мы видим много представительниц прекрасного пола с направлений истории, маркетинга, рекламы etc.

Может быть дело в том, что мальчики технари, а девочки - гуманитарии?

Выдвигаем гипотезу:



Представители мужского пола более способны к точным наукам, в то время как девушки лучше в гуманитарных предметах

---

### Находим информацию

Тут совсем просто. Заходим на <https://www.kaggle.com> и ищем интересующий нас датасет



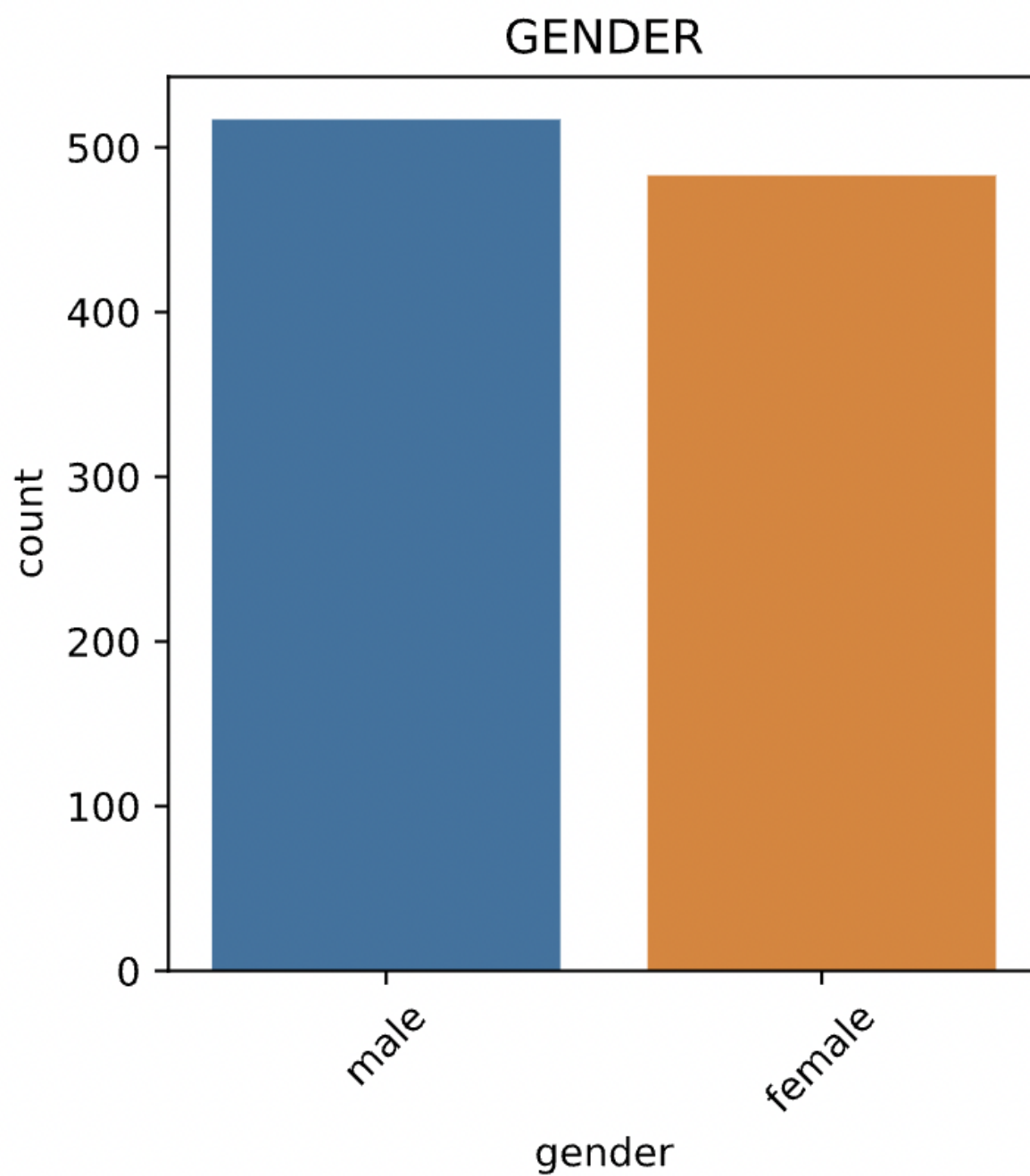
Вот то, что нам нужно

Датасет с результатами экзаменов по `reading`, `writing` и `maths` для разных студентов

---

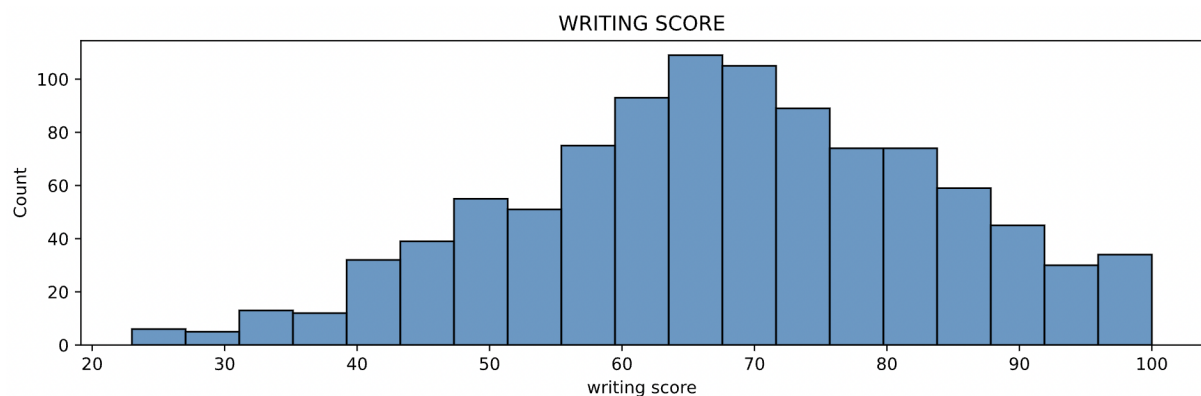
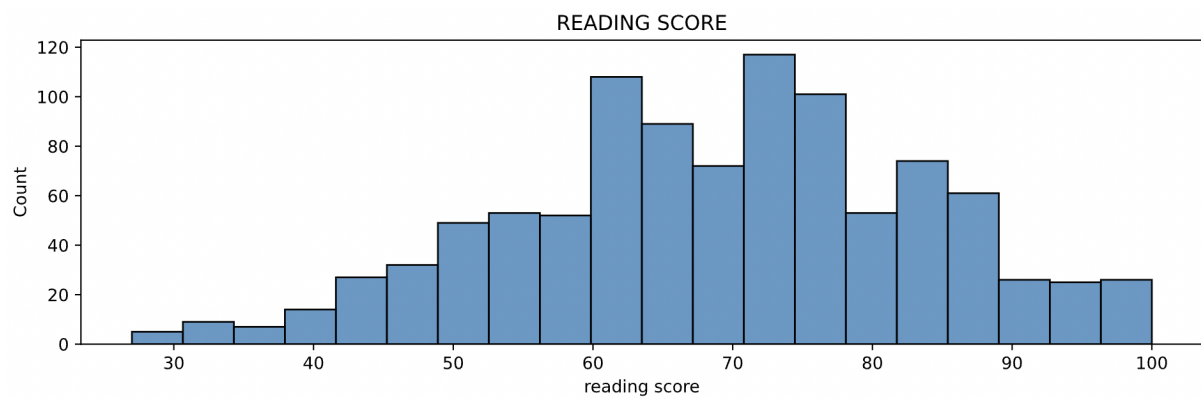
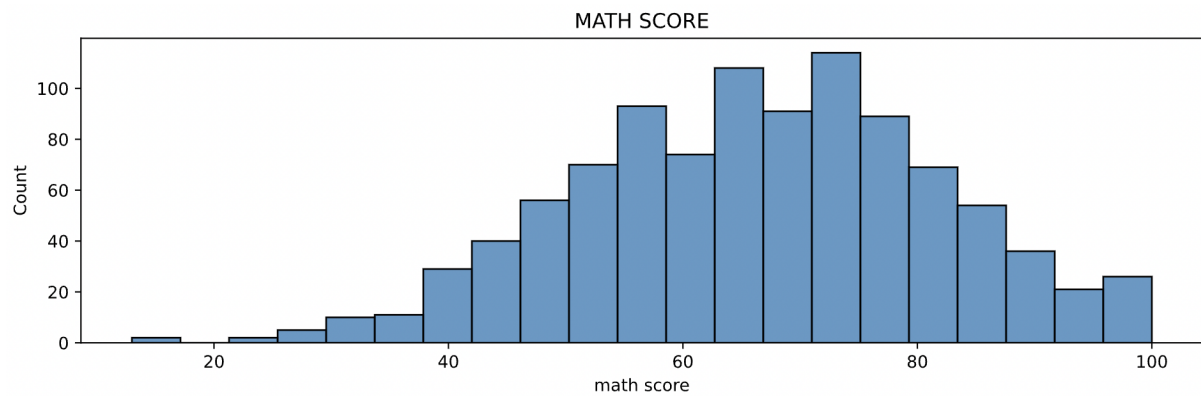
**Изучаем, рисуем вникаем**

Для начала убедимся, что все честно - девочек и мальчиков поровну:



Теперь посмотрим, на различные зависимости

Вот как распределились баллы по разным предметам:



Нормальное распределение на глаза. Значит, данные вполне из реальной жизни, им можно доверять (это не какая-то особенная школа с гендерным распределением по предметам)

Посмотрим на корреляцию между баллами по разным предметам

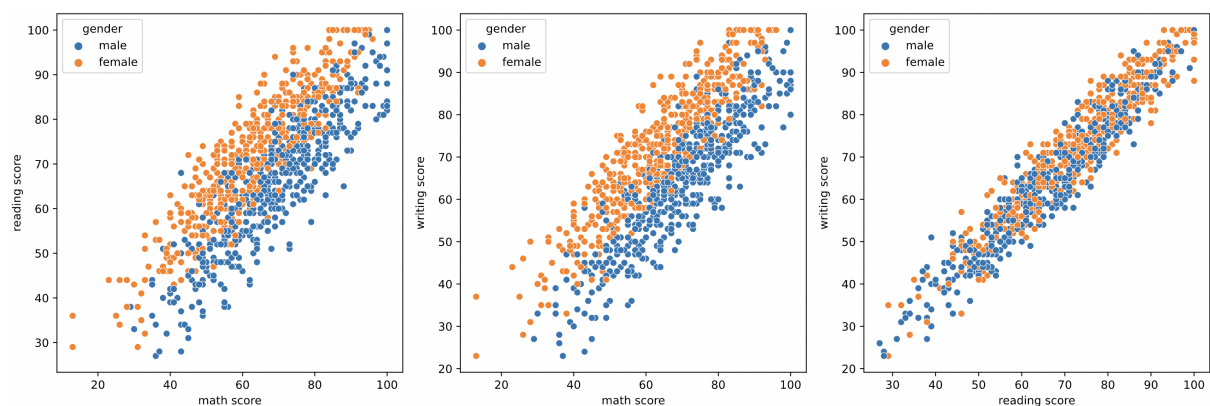
	math score	reading score	writing score
math score	1.000000	0.819398	0.805944
reading score	0.819398	1.000000	0.954274
writing score	0.805944	0.954274	1.000000

Отсюда понятно, что чаще всего студентам хорошо даются либо математика, либо чтение и письмо

### Ищем связи

Остается вопрос, *есть ли связь между предметом и полом студента?*

Посмотрим на график зависимости

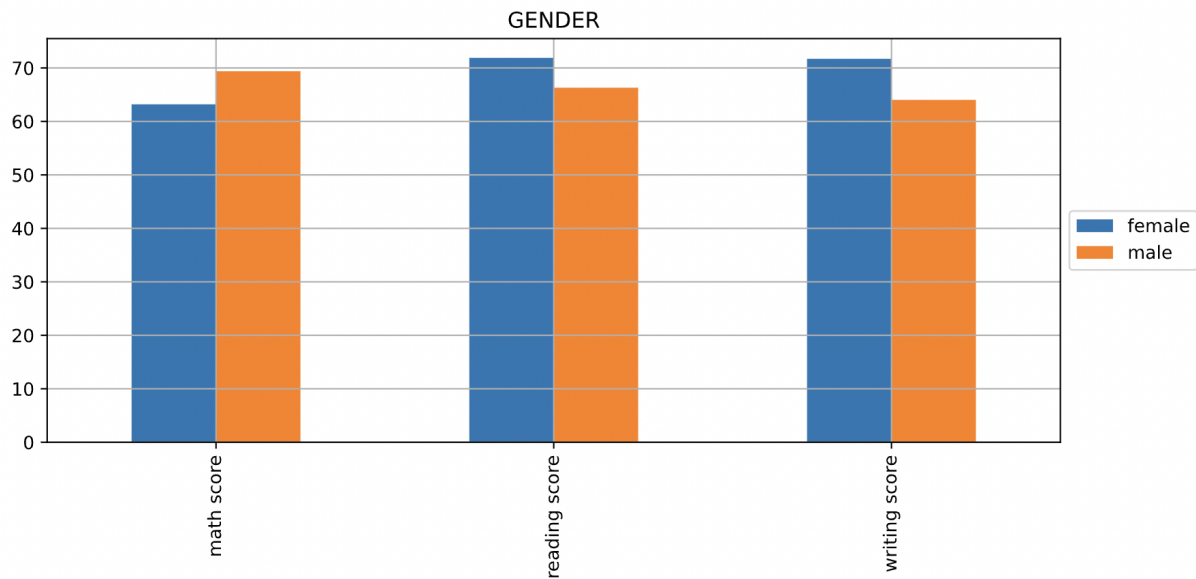


Прекрасно видно четкое разделение **мальчиков** и **девочек**

Простыми словами, чтобы не вчитываться - чем правее, тем лучше с математикой, чем выше - с гуманитарными дисциплинами

Таким образом, видно, что мальчики преуспевают в математике больше, чем девочки и больше, чем в гуманитарных науках. Аналогично и в обратную сторону

Если мы просто посмотрим на средний балл студентов, то картина будет предсказуемая:



COURSE: MATH

Average for female is 63.2.

Average for male is 69.38.

There is a significant effect of gender.

COURSE: READING

Average for female is 71.89.

Average for male is 66.31.

There is a significant effect of gender.

COURSE: WRITING

Average for female is 71.71.

Average for male is 64.03.

There is a significant effect of gender.



Ура!

Получилось доказать с помощью обычной статистики нашу гипотезу. Более того, даже показали, что если человек хорошо справляется с одной гуманитарной наукой, то и другая ему будет даваться хорошо.

### Показываем всем

Все красивые картинки выше (и не только, на самом деле их намного больше) это не просто рисунки, а графики.


А графики мы получаем с помощью кода, который расположен в открытом доступе и доступен к просмотру и комментариям.

## Результаты

Код тут:

HSE-ID-Course/exams.ipynb at main · heartsker/HSE-ID-Course

Contribute to heartsker/HSE-ID-Course development by creating an account on GitHub.

 <https://github.com/heartsker/HSE-ID-Course/blob/main/HW-3/exams.ipynb>

heartsker/**HSE-ID-Course**



1 Contributor 0 Issues 0 Stars 0 Forks

По всем вопросам: @heartsker