

Задание 5

Тема: Проверка статистических гипотез.

Для сдачи: 4 вопроса + 4 учебные задачи

Ограничения: 5 человек на 1 вопрос/задачу

Вопросы:

1. Что такое нулевая гипотеза? Если вы пытаетесь найти новую частицу, то что будет нулевой гипотезой в вашем эксперименте?
2. Что показывает уровень значимости для критической области статистики?
3. Приведите несколько статистических критериев, когда их следует использовать и при каких условиях?
4. В нейробиологии очень любили проверку гипотез. Она позволяла определить, какие части мозга на что реагируют. В результате проверяется каждая точка 2D или 3D изображения мозга. Однажды команда исследователей воспроизвела один из типичных дизайнов нейронаучных экспериментов, в котором испытуемому последовательно и много раз демонстрируются похожие стимулы, затем активность его мозга в ответ на эти стимулы сравнивается с активностью его мозга в состоянии покоя. Роль испытуемого в этом эксперименте играл мертвый лосось. В качестве стимула ему показывали картинки с изображениями людей в различных социальных ситуациях. Задача поиска областей, которые реагируют на этот стимул, была решена успешно. В мозге лосося были выявлены области, в которых активность значимо изменилась. Как такое могло произойти?
5. Пусть ваша лаборатория разработала новый метод протонной терапии. Вы провели ряд экспериментов, чтобы выяснить, что ваш метод действительно лучше, и получили p -value равно 0.04. Как бы вы оценили достоверность результата?

Задачи:

Учебные:

1. Muriel Bristol утверждала, что может сказать, чай или молоко было добавлено в чашку сначала. Ronald Fisher решил провести эксперимент: пред оставяется 8 чашек, в 4 наливают сначала молоко, в других 4 - чай; нужно определить 4 чашки одного типа. При каком результате можно будет сказать, что такая способность действительно есть? С чем могут возникнуть проблемы?
2. В 2020 году средний балл ЕГЭ по физике составил 54.5 балла, экзамен сдали 139 574 школьника. В 2019 году средний балл - 54,4, а сдавало 142 607 школьника. Является ли это увеличение среднего балла статистически значимым? (Прим. распределение баллов нормальное со среднеквадратичным разбросом - 6,1 балл; для подсчета статистической значимости можно использовать критерий Стьюдента)
3. Пусть эксперимент состоит в том, что вы считываете количество событий в каждом бине. Для этого возьмите полином какой степени и сгенерируйте на основе его данные (какое распределение стоит выбрать?). В качестве фитирующей функции возьмите полином той же степени и большей или меньшей, каждый полином в таком случае является гипотезой. Проверьте статистическую значимость гипотез.
4. В классической пачке m&m's 40 конфет, среди них в среднем красных - 13%, желтых - 14%, оранжевых - 21%, зеленых - 20%, коричневых - 12% и синих - 21%. Проверьте гипотезу, что распределение цветов конфет равномерное.
5. Монетку подбросили 1000 раз, 550 раз выпал орел. Как вы думаете, монетка честная? Почему?
6. Когда быстрые частицы проходят через диффузионную камеру, то ионизация молекул приводит к образованию капелек, которые выглядят, как треки. Предполагается, что вероятность образования капли на единицу длины трека постоянная и пропорциональная квадрату заряда частицы. Существование частицы с зарядом меньше единицы, "кварка", можно установить, если обнаружить трек, число капель которого на единицу длины было бы значительно меньше среднего. Из 55 000 треков со средним числом капель 229 был обнаружен трек только со 110 каплями. Посчитайте вероятность, что частица с единичным зарядом образует только 110 (или меньше) капель. Затем физика эксперимента была подвергнута сомнению, было указано, что в среднем рождается 4 капельки в каждом акте рассеяния. Считая, что рождается точно 4 капли, пересчитайте вероятность. Чувствителен ли результат к верности предположений?

Научные:

1. В понедельник утром после открытия бирж согласно экономической теории у трейдеров имеется максимум информации и максимум неопределенности, это должно выражаться в сплеске [спреда](#) в первые часы. Используйте [данные](#) четырех криптобирж, чтобы проверить гипотезу. Постройте также распределение разницы максимальных и минимальных цен и проверьте его на нормальность. Можно ли по QQ plot сказать, какое распределение лучше описывает. (reserved)