

Правила оформления и сдачи домашних заданий №4 и №5.

Решение задач №4 и №5 нужно сопровождать короткой запиской — текстом, написанном на естественном языке, из которого посторонний читатель смог бы понять, что вы делали и что получили.

За каждую задачу ставятся две отметки: за решение задачи и за устный ответ по темам курса, связанным с задачей. Вес самого решения — 40%, вес устного ответа — 60%. Отметки двух видов: зачтено/не зачтено. Если задача решена полностью и без ошибок, ставится оценка «зачтено», иначе — «не зачтено». Сдавать каждую задачу можно сколько угодно раз, но лично семинаристу или лектору — по почте решения присылать не надо. Выполнять расчёты можно в каких угодно программах.

Если задача сдаётся после указанного срока, оценка будет ставиться со штрафом (75% от полного балла).

Время от времени на устной защите возникают несогласия: например, студент считает, что даёт правильное определение, а преподаватель думает иначе. Мы будем признавать правоту студента, если тот укажет своё определение в какой-либо книге из списка рекомендованной литературы по этой теме — этот список есть в программе курса.

Задача №4.

Файлы «youtube_1.csv» ... «youtube_10.csv» содержат следующие сведения о видеороликах на YouTube (по сто роликов на каждый вариант):

n — номер наблюдения,

id — идентификатор ролика,

framerate — число кадров в секунду,

frames — общее число кадров в видео,

bitrate — битрейт, Кбит/сек.

duration — продолжительность, сек.

size — размер видеофайла, байт.

Для признаков **framerate**, **frames**, **bitrate**, **duration** и **size** рассчитайте две корреляционные матрицы — на основании коэффициентов Пирсона и Спирмена. Оцените значимость каждого коэффициента (проверьте гипотезу об отсутствии корреляции) и представьте полученные результаты в виде таблицы:

Коэффициенты корреляции Пирсона.

	framerate	frames	bitrate	duration	size
framerate	1.00	0.08	−0.02	0.04	0.02
frames	0.08	1.00	0.12	0.45**	0.29*
bitrate	−0.02	0.12	1.00	−0.03	0.72***
duration	0.04	0.45**	−0.03	1.00	0.36**
size	0.02	0.29*	0.72***	0.36**	1.00

* — коэффициент значим на уровне 5%,

** — коэффициент значим на уровне 1%,

*** — коэффициент значим на уровне 0.1%.

Коэффициенты, не отмеченные звёздочками, незначимы (нет оснований отвергнуть гипотезу об отсутствии корреляции на уровне 5%).

Числа в приведённой таблице я проставил с потолка, смысла в них нет. Это просто пример. Таблица должна быть оформлена ИМЕННО ТАК.

Сравните коэффициенты Пирсона и Спирмена, обратите внимание на случаи, когда два этих коэффициента существенно расходятся, если такие есть. Что такое «существенно», решайте сами. В случае существенного расхождения постройте диаграммы разброса для тех пар признаков, тесноту связи между которыми коэффициенты измеряют по-разному, и попытайтесь объяснить причину расхождения.

Если вы не видите никаких существенных расхождений между двумя матрицами, просто постройте диаграмму рассеяния для случая, где разность коэффициентов Пирсона и Спирмена наибольшая.

Срок сдачи: до 22 мая включительно.