

А.В. Браилов. Математическая статистика. Правила оформления и подсчет баллов за курсовую работу — 2016

Работа должна содержать следующие разделы:

Содержание (**не** Оглавление!)

1. Введение
2. Предварительный анализ данных
3. Теоретическая справка по проверке гипотез
4. Проверка гипотезы для модельных данных
5. Выбор альтернативной гипотезы и оценка мощности критерия
6. Проверка гипотез для реальных данных
7. Заключение
8. Литература
9. Приложения 1, 2 (1 — характеристики компьютера, 2 — коды программ)
10. Приложение 3 (Список файлов)

Разделы **1 — 6, 8 — 10** оцениваются от 0 до 7 баллов. Заключение при наличии новизны дает до 10 баллов. Оформление работы оценивается от 0 до 7 баллов (дополнительно). Так как каждый раздел **оценивается отдельно, необходимо соблюдать указанный порядок разделов и их нумерацию**. Нарушение этого правила снижает до 0 оценку **оформления**.

Работа сдается в **два этапа**: к промежуточному контролю и в конце семестра с защитой. Своевременное выполнение этих этапов дает дополнительно до 20 баллов:

- введение и предварительный анализ данных с титульным листом и в папке (к **промежуточному контролю**) — до 10 (0, 2, 4, ...);
- сдача всей работы в **срок** — еще до 10 баллов.

Замечание. Для промежуточного контроля оценка за введение и предварительный анализ данных на сайте МК выставляется от 0 до 15 баллов: 0, 3, 6, ..., 15.

Своевременность сдачи работы в конце семестра оценивается следующим образом:

- 10 баллов, если работа сдана в срок;
- 8 — при опоздании на одну неделю;
- 5 — на две недели;
- 0 баллов — при опоздании на три недели и больше.

Датой сдачи работы считается дата, **более поздняя** из двух:

- даты получения электронной версии научным руководителем;
- даты поступления бумажной версии научному руководителю (на кафедру, если руководитель не ведет практические занятия).

Работа принимается только при наличии идентичной электронной версии. Повторная сдача курсовой работы, как правило, **не допускается**. Если основанием для повторной сдачи является неудовлетворительная оценка первого варианта, второй вариант оценивается не более 69 баллов. В случае опоздания более трех недель работа принимается только при наличии объяснительной записки с резолюцией декана.

Требования к содержанию разделов 1 — 10

1. Введение должно занимать не более одной страницы и содержать разделы:

1.1. Уточнение темы. Если в теме не указан основной критерий, необходимо его выбрать; расшифровать, что значит «определенного», ...

1.2. Описание исследуемых данных (в том числе **индекса**).

1.3. Число проверок основной гипотезы. Описание соответствующих выборок (с указанием их точного объема), к которым будет применен основной критерий.

1.4. Планируемая новизна.

2. Предварительный анализ данных. Необходимо описать методику выбора исследуемых ценных бумаг, дать список соответствующих компаний и тикеров, указать источники списков компаний и котировок акций. Если список компаний составляется по данным из Википедии, необходимо указать дату последнего обновления соответствующей страницы. Эта дата должна подтверждать актуальность списка.

Предварительный анализ данных содержит: таблицу числа торговых дней (по годам и акциям), таблицы максимальных (вверх и вниз, **две** таблицы) дневных относительных скачков цен (по годам и акциям), графики цен для акций с максимальным однодневным снижением (повышением, отдельный график) цены. **Для экстремальных скачков** следует указать соответствующие **номера** торговых дней и **указать** их на графиках.

Необходимо перечислить названия используемых полей (например, Adj Close, VOL) единицы измерения и смысл полей. Если некоторые временные ряды вызывают сомнения, список исследуемых бумаг сокращается (с объяснением причины). Для окончательного списка акций производится сокращение данных: неиспользуемые поля и данные, выходящие за пределы окончательного диапазона лет, удаляются. При этом данные за последний полный календарный год должны быть представлены. Ссылки на Приложение 2 не допускаются!

3. Теоретическая справка по проверке гипотез. Содержит описание основного статистического критерия и вспомогательного критерия по проверке равномерности распределения P -значения основного критерия. Оценка снижается как за недостающие, так и за **лишние** сведения. Поощряется применение обозначений, используемых в наших курсах теории вероятностей и математической статистики. **Все нестандартные**

обозначения должны быть определены до их первого появления! Необходимо указать, по каким источникам написана теоретическая справка.

4. Проверка гипотез для модельных данных. На основе теоретической справки описывается программа, создающая методом Монте-Карло таблицу 999 квантилей распределения статистики основного критерия при верной нулевой гипотезе (статистика вычисляется не менее 10000 раз) при объемах выборки, **точно соответствующих** конкретным выборкам из п.6. Распечатываются только квантили уровней 0.1, 0.2, ..., 0.9.

Объясняется, каким образом, была получена гистограмма P -значений для модельных данных (1000 P -значений). Производится проверка равномерности распределения P -значения на отрезке $[0,1]$ по критерию Колмогорова (равномерность должна подтверждаться). Ссылки на Приложение 2 **не допускаются**, так как такие ссылки создают иллюзию объяснения появления и смысла соответствующих рисунков и таблиц.

5. Выбор альтернативной гипотезы и оценка мощности критерия. Выбираются три альтернативные гипотезы. Единственное ограничение на выбор альтернативных гипотез — они не должны слишком сильно отличаться от нулевой гипотезы. Мощность критерия приближенно оценивается на основе 1000-кратной проверки нулевой гипотезы (с уровнем значимости 0.05) при верных альтернативных гипотезах. Объемы искусственных выборок должен **точно** соответствовать объемам реальных исследуемых выборок. Ссылки на Приложение 2 не допускаются!

6. Проверка гипотез для реальных данных. Одним из главных результатов работы является таблица P -значений основного критерия (по периодам и компаниям), которая обязательно иллюстрируется гистограммой (одна общая гистограмма для всех периодов и тикеров). Вычисляется доля проверок, в которых гипотеза принималась при 5% и 1% уровнями значимости. Если основная гипотеза систематически отвергается даже при 1% уровне значимости, необходимо сократить исследуемые периоды времени (например, анализ полного календарного года можно заменить исследованием данных за второе полугодие). Вычисляются медианные P -значения: а) по периодам и б) по компаниям. *Для четырех экстремальных медиан P -значений (2 — по периодам и 2 — по компаниям) следует попытаться найти экономическое объяснение.* Как и в других разделах, ссылки на Приложение 2 не допускаются! Вместо ссылок должны быть описаны действия, которые привели к полученным результатам. Проверка равномерности распределения реальных P -значений на отрезке $[0,1]$ не производится.

7. Заключение содержит разделы:

- **7.1. Новизна программ** (рисунков, таблиц).
- **7.2. Новизна статистических эффектов.**
- **7.3. Сравнение полученных результатов** с результатами наиболее близких по теме (или данным) работ прошлых лет.

Для новых тем необходимо указать наиболее близкие старые темы и сопоставить полученные результаты, если такое сопоставление имеет смысл. Для старых тем такое сопоставление выполняется в обязательном порядке. **Особо отмечаются** результаты, которые в предыдущих курсовых работах не были получены или им противоречат.

8. Литература. Оценивается по числу (новых) правильно оформленных ссылок на журнальные статьи и книги по теме курсовой работы. Цитированные во введении курсовые работы включаются в список литературы. Ссылки на Интернет не учитываются при подсчете баллов, однако источники данных и списков тикеров указываются.

9. Приложения 1 и 2. Приложение 1 содержит характеристики компьютера, на котором будут выполняться основные вычисления: тип процессора (например, Intel Celeron E3300), тактовая частота и частота системной шины, объем кэш-памяти второго уровня (L2). Это приложение включает также список программ, работающих более 10 секунд.

Приложение 2 содержит коды всех программ. Программы приводятся в порядке появления создаваемых ими таблиц и рисунков. Каждая программа должна **начинаться с комментария, содержащего номер рисунка (таблицы) и краткого пояснения, повторяющего текст под рисунком (над таблицей).**

Если некоторая программа во многом **повторяет предыдущую программу**, то в курсовую работу переносятся **только первые строки**, в которых имеются отличия. Вместо остальных строк ставится многоточие.

Учитывается оформление (отступы, расположение скобок, наличие необходимых комментариев). При переносе кода программы в MS Word длинные строки кода либо обрезаются (что недопустимо), либо переносятся на следующую строчку, что некрасиво. Чтобы избежать подобных негативных явлений, сразу пишите программы с короткими строками или уменьшайте кегль шрифта кода.

Если программа новая, автор указывает свою фамилию и год разработки в комментарии в начале программы. Если программа работает долго (более 10 секунд), указывается приблизительное время работы программы. Если автор программы не известен, указывается источник (курсовая работа). Если программа редактировалась, редактор указывает свою фамилию и год редакции после фамилий автора. В Приложении 2 **измененный код выделяется полужирным шрифтом.**

10. Приложение 3. Список файлов. Это приложение содержит перечень файлов, передаваемых *до* сдачи бумажной версии курсовой работы. Название файла программы должно **начинаться с номера создаваемой этой программой таблицы (Таб.2.1.Максимальные скачки по годам...)** или рисунка (**Рис.2.1.График цены...**).

Отсюда следует, что каждому рисунку соответствует ровно **одна программа**. Каждой таблице соответствует ровно **одна программа**. Таким образом, число программ

равно суммарному количеству рисунков и программ. Если две программы мало отличаются, эти отличия должны быть сосредоточены в **первых строчках!**

Работа сдается: сначала в электронном виде, затем — в бумажном виде. Бумажная версия приносится на практическое занятие (в **особых** случаях — на кафедру) в тонкой папке с прозрачным верхом. Идентичная электронная версия присылается по почте в виде ZIP-архива. **Архивы формата RAR не принимаются!**

Архив называется **Фамилия.zip** и содержит папку **Фамилия**. Папка **Фамилия** содержит: а) курсовую работу **Фамилия.docx**; б) папку **Prog** с используемыми **программами и сокращенными данными**. Исходные данные, **Prog0**, научному руководителю **не пересылаются** (но приносятся **на защиту**).

ZIP-архив должен иметь указанную структуру и размер не более 1 МБ. Исключения (по размеру) допускаются только по согласованию с преподавателем. **При неполном наличии заархивированных файлов или их несоответствии бумажной версии курсовой работы оценка за Приложение 3 снижается до 0.**

Требования к оформлению работы

Папка (тонкая, с прозрачным верхом, страницы сшиты), титульный лист (**Федеральное государственное образовательное бюджетное учреждение высшего образования**), нумерация страниц (титульный лист считается), содержание, шрифты, выравнивание текста и формул. Формулы набираются с помощью MathType, Equation или как «формула Word 2007(или старше)». Оценивается вид рисунков и таблиц.

Все таблицы и рисунки должны быть пронумерованы (например, **Рис. 2.1,...**). Под рисунками должны быть содержательные (но не слишком большие) подписи, а **над** таблицами — надписи.

В ссылках на рисунки и таблицы используются сокращения «табл.» и «рис.» (первая буква — **строчная**, если это не начало предложения). В конце заголовков точка не ставится! После номера раздела, рисунка или таблицы, напротив, точка необходима.

Объем

Общий объем курсовой работы (вместе с приложениями) находится в пределах от 15 до 25 страниц. Работы, содержащие более 25 страниц, без предварительного разрешения не принимаются. Основанием для увеличения объема может быть только наличие интересных результатов!

Защита

На защите заслушивается короткий доклад-презентация (3 — 10 минут), проводится выборочная проверка работоспособности программ, задаются вопросы по курсовой работе и программам. В докладе отмечается наличие (отсутствие) новизны. При наличии новизны время доклада увеличивается до 10 минут. Новизна является основным содержанием доклада.

На защите предварительная оценка курсовой работы может корректироваться как в сторону увеличения (5 баллов за презентацию), так и в сторону уменьшения (за ответы на вопросы).

Все программы и данные записываются на компьютеры самим студентом до начала защиты.

Основные программы и сокращенные данные содержатся в папке Prog, исходные данные и программа создания сокращенных данных — в папке Prog0.

Форматирование основного текста

- шрифт Times New Roman (или Cambria — Calibri, если Word 2007 — 2010);
- кегль шрифта 13;
- междустрочный интервал полуторный;
- левое поле 2,5 см;
- правое поле 1 см;
- автоматическая расстановка переносов;
- выравнивание по ширине строки;
- отступ первой строки 0,8 см;
- ссылки на литературу указываются в квадратных скобках: [1], [2, с.32–35] и т.д.

Форматирование Приложений 1 — 3

- шрифт Courier New;
- кегль шрифта 11;
- междустрочный интервал 1,2;
- выравнивание по левому краю.

Заключительные замечания

- Не используйте неоправданно большую точность в таблицах.
- Чтобы поместить таблицу на одной странице уменьшайте кегль и интервал.
- Запрещается цитировать курсовые работы прошлых лет по теоретическим вопросам (по обнаруженным эффектам — можно и нужно).
- Точность в таблицах P -значений — 10^{-3} .
- Все новые программы и добавленные/измененные строки старых программ включаются в основные разделы работы вместе с обязательными пояснениями.
- На первой странице не должно быть номера.
- Работа проходит компьютерную проверку на плагиат.