**ЗАДАНИЯ К ЗАЧЁТУ 2020. ВАРИАНТ 1**

**Выполняют те, чья фамилия начинается на буквы А-И**

***Инструкция****. В таблице представлены гипотетические данные по 105 особям Лемуров (Lemur sp.) (К биологии каких-либо реальных видов они отношения не имеют).*

***Нужно выполнить все задания, из которых 4 или 5*** *–* ***в R****,* ***остальные в Statistica*** *(по выбору, но задания 5 и 6 точно в R)*

*Ответ на любое задание следует начинать с графического анализа данных (гистограммы), проверки соответствия данных условиям нормального распределения (кроме заданий 1, 2, 3, 5). (Переменные с линейными промерами - столбцы I-N – проверять на нормальное распределение не надо ни в каких заданиях, везде считать их нормальными). В случаях, когда переменные требуют трансформации, трансформируйте их (!). В многомерных методах не забудьте стандартизировать переменные. В одном-двух заданиях по выбору необходимо посчитать для любой переменной описательную статистику (Среднее, показатели разброса и пр.) в модуле описательной статистики.*

*Для работы в* ***Statistica*** *следует сохранить таблицу в Statistica (.STA) и назвать её своей фамилией (Иванов). Некоторые задания предполагают добавление* ***новых столбцов*** *с данными в исходную таблицу. Все результаты статистических тестов в Statistica (в том числе промежуточные) сохраняйте, пожалуйста, в Рабочую книгу (.STW), названную вашей фамилией.*

*Для работы в* ***R*** *рекомендуется импортировать в R таблицу так, чтобы столбцы для предполагаемого анализа не содержали пустых ячеек; в R таблицу можно называть произвольно. В R сохраняйте, пожалуйста, Вывод в виде файла TXT, названного Вашей фамилией, в который обязательно копируйте скрипт (без результатов, по образцу моих текстовых файлов). Можно сделать несколько файлов с выводом, но название должно содержать фамилию.*

*Ответ на задания нужно набить* ***в документ Word*** *и назвать его своей фамилией с названием института. В каждом задании нужно как в статье, сначала охарактеризовать полученный результат словами, а потом в скобках привести название критерия, статистику критерия и уровень значимости p. Например: «масса лемуров была достоверно связана с полом: самцы весили больше, чем самки (t-критерий Стьюдента для независимых выборок, t=3.6, p=0.02)». Копировать картинки (гистограммы и пр.) в файл Word не нужно.*

*Мне на почту* [*ninavasilieva@gmail.com*](mailto:ninavasilieva@gmail.com) *нужно обязательно выслать* ***4 файла****: .STA (табличка), .STW (рабочая книга), .TXT (вывод и скрипт в R) .DOC или .DOCX (сам текст ответов).*

*Срок выполнения –* ***строго до 17 мая****.*

1. Различался ли уровень социальной активности лемуров в разные сезоны года *(летом-осенью; не забыть про трансформацию данных; Столбцы P и Q)*?
2. Соответствует ли соотношение всех самцов и самок соотношению 1:1? *(столбец B; речь идёт о сравнении наблюдаемых частот с теоретическими; придётся подсчитать, сколько всего самцов и самок, и какими будут ожидаемые частоты для N=105)*.
3. Различается ли соотношение полов в разных возрастных группах лемуров (*столбцы B и C; в Статистике это раздел Tables and Banners)*?
4. С помощью дисперсионного анализа проанализировать, как различается длина тела (*столбец L*) у особей разного пола и возраста *(Столбцы B и C)*; обязательно учесть возможное взаимодействие факторов и провести пост-хок тест.
5. С помощью перестановочного теста (пакет coin) сравнить массу тела у детёнышей разного пола (*Столбцы S и U*)
6. Построить линейную модель зависимости массы тела (*столбец E*) от пола, возраста и длины тела (*столбцы B, C, L*), добавив номер семьи (*столбец D*) как случайный фактор: произвести 7 шагов анализа линейной модели (лекция 8). В том числе, выбрать лучшую модель с помощью критерия Акаике (напр., пакет MuMIn); оценить р-значение для *возраста* с помощью теста отношения правдоподобий (функция *anova*); провести тест Тьюки для возраста в пакете emmeans*.*
7. Построить модель того, как на вероятность принести потомство (*столбец R*) влияют масса тела и социальный ранг *(Столбцы E и H).*
8. На основе зависимости массы тела от длины тела (*Столбцы E и L*) получить переменную «упитанность»: провести анализ остатков (residual analysis как составляющая регрессионного анализа) и вставить в таблицу остатки (residuals) в качестве новой переменной.
9. провести анализ главных компонент над переменными, описывающими линейные размеры лемуров *(Столбцы I-N)*. Предварительно обязательно построить матрицу корреляций. Выделить два фактора из 6-и исходных переменных.
10. На основе линейных размеров лемуров (*Столбцы I-N*) провести дискриминантный анализ четырёх возрастных групп (*столбец C*): указать, какие переменные внесли вклад в разделение групп, получить картинку в пространстве дискриминантных функций (в Perform canonical analysis) и указать процент правильного причисления особей для каждой группы (указан в Classification matrix).

*Примечания:*

1. частотные критерии можно рассчитывать через Basic Statistics – Tables& Banners (не забудьте поставить галочки соответствующим тестам во вкладке Options и выбрать detailed two-way tables).

2. некоторые результаты могут быть **недостоверны**, это **нормально**!