Любые вычисления (даже промежуточные) необходимо округлять не менее, чем до двух знаков после запятой.

для расчета стандартного отклонения Excel предлагает несколько функций,

но лучше всего использовать эту:

=СТАНДОТКЛОН.В()

для расчета логарифма переменной (даже в русскоязычном Excel) нужно

использовать следующую функцию:

=LN()

для расчета квантилей стандартного нормального распределения,

хи-квадрат распределения, распределения Стьюдента и распределения

Фишера нужно использовать следующие функции соответственно:

=НОРМ.СТ.ОБР()

=ХИ2.ОБР()

=СТЬЮДЕНТ.ОБР()

=F.ОБР()

для расчета вероятностей вида P{X < x} для номарльного распределения,

хи-квадрат распределения, распределения Стьюдента и распределения

Фишера нужно использовать следующие функции соответственно:

=НОРМ.РАСПР()

=ХИ2.РАСПР()

=СТЬЮДЕНТ.РАСПР()

=ХИ2

=F.РАСПР()

будьте внимательны с последним аргументом каждой из этих функций:

его нужно выставить на ИСТИНА, т.е. использовать интегральную функцию

распределения

для генерации случайных чисел из равномерного на отрезке [0, 1],

экспоненциального с параметром lambda и нормального распределений

нужно использовать следующие функции соответственно:

=СЛЧИС()

=LN(СЛЧИС())/ lambda

=НОРМ.ОБР(СЛЧИС())

2

на самом деле, лучше всего генерировать нормальные случайные

величины, используя надстройку Excel, которая называется "Анализ

данных", пункт "Генерация случайных величин"

Прикладная

Егор выбирает себе квартиру в одном из двух районов города М. Надо помочь Егору, проанализировав данные по квартирам в соответствующих районах. Сами данные представляют собой четыре переменные (четыре показателя): цены квартир (млн.руб.) (flat_price_region_1 и flat_price_region_2) и их жилая площадь (м.кв.) (living_space_region_1 и living_space_region_2) в каждом из двух районов. Выполните следующие задания и подготовьте файлы к отправке.

Часть 1.

1. Рассчитайте дескриптивные (описательные) статистики для всех четырех переменных: минимальное значение, максимальное значение, среднее значение, стандартное отклонение, медианное значение. В текстовый файл запишите, что собой представляют переменные, указав при этом значения всех дескриптивных статистик. Для чего, по вашему мнению, рассчитывают дескриптивные статистики? Для каждой переменной сравните ее среднее значение и медиану. Объясните, как вы понимаете различие/сходство между ними. Что вы можете сказать о стоимости квартир в двух районах на основе дескриптивных статистик?

2. Диаграмма рассеяния (scatter plot) - постройте ее для жилой площади квартир и их стоимости для каждого района в отдельности. Рассчитайте выборочную корреляцию между жилой площадью квартир и их стоимостью для каждого района в отдельности. Вставьте в текстовый файл и графики диаграмм рассеяния, и значения выборочных коэффициентов корреляции, указав формулу, по которым они рассчитываются. Объясните, как соотносятся между собой диаграмма рассеяния и значение выборочной корреляции.

3. Предположим, имеются две зависимые переменные: X_1 и X_2..

(a) Может ли выборочная корреляция между этими переменными оказаться близкой к 0? Почему?

(b) Предположим теперь, что истинное значение корреляции между переменными X_1 и X_2 равно 0. Может ли выборочная корреляции оказаться отличной от 0? Почему?

Часть 2. Статистический анализ данных.. Тестирование гипотез, построение доверительных интервалов.

1. Постройте гистограмму распределения стоимости квартир во втором районе. Какое из известных вам распределений она напоминает? Почему вы так решили? Поместите гистограмму в текстовый файл, указав (аргументированно) распределение, на которое

она похоже (указывать значение параметров этого распределения не нужно). В дальнейшем считайте, что распределение данной переменной именно такое, которое вы предположили.

2. Оцените аналитически неизвестный параметр распределения стоимости квартир во втором районе при помощи метода максимального правдоподобия. В текстовый файл запишите функцию правдоподобия, логарифмическую функцию правдоподобия, первую производную, оценку параметра, информацию Фишера и покажите выполнение условия второго порядка.

3. При помощи теста отношения правдоподобия (LR-теста) проверьте гипотезу о том, что истинное значение параметра распределения стоимости квартир во втором районе равно 0.4. В текстовый файл запишите основную и альтернативную гипотезы, укажите выражение для расчетной статистики, ее распределение при верной основной гипотезе, приведите значение p-value и проинтерпретируйте результат, используя заданный для вашего варианта уровень значимости.

4. Используя заданный уровень значимости, при помощи теста Колмогорова проверьте гипотезу о том, переменная стоимости квартир во втором районе действительно получена из предполагаемого вами распределения с параметром 04. В текстовый файл запишите основную и альтернативную гипотезы, укажите выражение для расчетной статистики и ее распределение при верной основной гипотезе. Приведите результат тестирования гипотезы.

5. Используя заданный уровень значимости, постройте доверительный интервал для вероятности того, что стоимость случайной квартиры во втором районе будет не больше 5 или не меньше 15. В текстовом файле укажите общий вид используемого доверительного интервала, а также его реализацию. Как интерпретировать ситуацию, когда реализация доверительного интервала для вероятности оказывается вне границ интервала $[0, 1]$?

6. Используя заданный уровень значимости, постройте доверительный интервал для математического ожидания стоимости квартир в первом районе. В текстовый файл запишите общий вид используемого доверительного интервала, а также его реализацию.

7. Проверьте гипотезу о том, что математические ожидания жилых площадей квартир в двух районах совпадают, против альтернативы, что математическое ожидание жилой площади квартир в первом районе больше математического ожидания жилой площади квартир во втором районе. Рассчитайте p-value соответствующего теста. В текстовый файл запишите основную и альтернативную гипотезы, укажите выражение для расчетной статистики, ее распределение при верной основной гипотезе, приведите

значение p-value и проинтерпретируйте результат, используя заданный для вашего варианта уровень значимости.

Также рассчитайте:

Если вам кажется, что в вашем варианте распределение рассматриваемой переменной нормальное, оцените при помощи метода максимального правдоподобия только параметр μ, считая $\sigma_2 = 3$.

5.

(a) вероятность ошибки первого рода;

(b) вероятность ошибки второго рода, предполагая, что альтернативная гипотеза является простой и утверждает, что разница между значениями для математических ожиданий жилых площадей квартир в первом и во втором районах равна 20;

(c) мощность критерия тестируемой гипотезы.

8. Проверьте гипотезу о том, что доля квартир первого района, чья стоимость превышает 5, совпадает с долей квартир второго района, чья стоимость также превышает 5 млн.руб. против альтернативы, что эти доли не совпадают. Для удобства вы можете создать дополнительные переменные. Рассчитайте p-value данного теста. В текстовый файл запишите основную и альтернативную гипотезы, укажите выражение для расчетной статистики, ее распределение при верной основной гипотезе, приведите значение p-value и проинтерпретируйте результат, используя заданный для вашего варианта уровень значимости.

Часть 3. Теоретические нюансы статистики. В столбце .part_3. вашего набора данных представлена реализация выборки из распределения Пуассона с неизвестным параметром $\lambda > 0$. Вам необходимо построить 90%-ый доверительный интервал для функции $g(\lambda) = 2|\lambda - 5| - 4 = 0$. Обратите внимание, что использовать, например, дельта-метод здесь не получится, функция $g(\lambda)$ не является дифференцируемой. Как можно построить доверительный интервал в данном случае? Реализуйте предложенный способ самостоятельно на компьютере и опишите реализованную процедуру.

Примечание. Для того, чтобы ответить на поставленный вопрос, ознакомьтесь с разделами . Overview. и .Introduction to the Bootstrap and Permutation Tests. в статье Hesterberg T. (2014). .What Teachers Should Know about the Bootstrap: Resampling in the Undergraduate Statistics Curriculum..