

GeekBrains, ML in Business

Lesson 8 Homework

Задание 1

Если мы рассчитали 95% доверительный интервал для среднего значения, то какие из следующих утверждений являются верными?

Выберите несколько вариантов и попытайтесь объяснить свой выбор:

1. Если многократно повторять эксперимент, то 95 % выборочных средних значений будут принадлежать рассчитанному нами доверительному интервалу.
2. Мы можем быть на 95% уверены, что среднее значение в генеральной совокупности принадлежит рассчитанному доверительному интервалу.
3. Если многократно повторять эксперимент, для каждой выборки рассчитывать свой доверительный интервал, то в 95 % случаев истинное среднее будет находиться внутри доверительного интервала.
4. Среднее значение в генеральной совокупности точно превышает нижнюю границу 95% доверительного интервала.
5. Среднее значение в генеральной совокупности точно принадлежит рассчитанному доверительному интервалу.

Решение Задания 1

1, 2.

Определение длительного интервала - по сути пункт 2. Также **при качественной и достаточно большой выборке** значения выборочных средних будут стремиться к значению среднего генеральной совокупности, как раз находясь в этом 95% доверительном интервале.

Задание 2

Если мы увеличиваем объем выборки в два раза (при условии, что показатель стандартного отклонения остается неизменным), то 95% доверительный интервал ...

Выберите один вариант из списка:

1. Стал более узким
2. Возможны оба варианта
3. Стал более широким

Решение Задания 2

1. Станет более узким.

Чисто практически, это происходит потому, что для нахождения доверительного интервала мы делим отклонение от среднего на \sqrt{n} .

Если смотреть на это концептуально, то при большем количестве наблюдений согласно вероятности мы можем быть более уверены в доверительном интервале при одинаковом стандартном отклонении. Раз мы более уверены в предыдущем 95% доверительном интервале, новый 95% доверительный интервал будет уже.

Задание 3

В центре 95% доверительного интервала, рассчитанного по выборочным значениям, находится ...

Выберите один вариант из списка:

1. Значение стандартной ошибки среднего
2. Выборочное среднее значение
3. Среднее значение генеральной совокупности

Решение Задания 3

1. Выборочное среднее значение.

Это следует из формулы и идеи выборочного интервала.

Задание 4

Часто на практике нулевая гипотеза отклоняется, и различия считаются статистически достоверными, если $p < 0,05$. Однако часто в статистике используется более жесткий критерий достоверности различий, например, при условии, что $p < 0,01$. Значение p -уровня значимости, которое выбирается, в качестве порога обозначается буквой α (альфа). Например, если исследователь решил, что $\alpha = 0,05$, то и нулевая гипотеза будет отклоняться при условии, что $p < 0,05$.

Если в определенной ситуации весьма рискованно отклонить нулевую гипотезу, когда она на самом деле верна, то лучше использовать показатель α равный ...

Выберите один вариант из списка:

1. 0,1
2. 0,001
3. 0,05
4. 0,5

Решение Задания 4

1. 0.001.

Раз рискованно отклонять гипотезу, мы хотим дать ей как можно больше вероятности быть верной. Только если вероятность совсем мала, мы хотим ее отклонить. Поэтому стоит взять самое маленькое значение - 0.001.

Но также надо сказать, что требуется знать примерный уровень риска, ведь 95% уверенность (0.05 альфа) тоже достаточно неплоха во многих ситуациях.

Задание 5

Данные некоторого исследования сообщают нам, что средний рост детей в 14 лет составляет 166 сантиметров. Однако это лишь выборочная оценка, и исследователи рассчитали 95% доверительный интервал, который составил [160, 172].

Укажите верные утверждения, выберите несколько вариантов и попытайтесь объяснить свой выбор:

1. У нас достаточно оснований отклонить нулевую гипотезу, что среднее в генеральной совокупности равняется 173
2. Вероятность того, что истинное среднее значение больше 172, составляет 0,01
3. У нас достаточно оснований отклонить нулевую гипотезу, что среднее в генеральной совокупности равняется 158.
4. Доверительный интервал не может иметь такие границы, т. к. выборочное стандартное отклонение равняется 10, следовательно доверительный интервал должен быть значительно шире.

Решение Задания 5

п.1. Верно. Так как у нас вероятность 95%, что генеральное среднее внутри доверительного интервала [160, 172]. 173 снаружи этого интервала, значит вероятность того, что это среднее (p-value) точно < 0.05 , значит гипотеза отвергается.

п.3. Верно. Аналогично пункту 1.

Пункт 2 неверен (вероятность будет 0.025), а в пункте 4 недостаточно информации, т.к. мы не знаем размер выборки.

Задание 6

Предположим, нулевой гипотезой вашего исследования являлось предположение, что конверсия в генеральной совокупности равняется 0.4. Вы получили $p = 0,12$ и не смогли отклонить нулевую гипотезу. Однако позже выяснилось, что конверсия в генеральной совокупности действительно равна 0.4. Как можно оценить результаты?

Какой вариант корректный по вашему мнению и почему:

1. Вы не совершали ни ошибку первого рода, ни ошибку второго рода.
2. Вы совершили ошибку первого рода
3. Вы совершили ошибку второго рода

Решение Задания 6

1. Верно. Мы не отклоняли гипотезу и это верно, значит ошибок нет.