МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение   
высшего образования

**«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Кафедра «Прикладная математика и информатика»

ТИПОВОЙ РАСЧЕТ

по дисциплине

«Теория вероятностей и математическая статистика»

Выполнил:

студент группы ИВТАПбд-21

Вершинин Д. В.

(вариант N = 4, V = 2)

Проверил:

Клячкин В. Н.

Ульяновск, 2017

**ЗАДАНИЕ 1. МЕТОДЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Вариант | N | m0 | σ | n | k | n1 |
| 4 | 660 | -6,5 | 2,2 | 48 | 9 | 18 |

**1.** Используя генератор случайных чисел пакета Анализ данных процессора электронных таблиц Excel, смоделировать генеральную совокупность из N нормально распределенных чисел с заданным математическим ожиданием m0 и стандартным отклонением σ.

Смоделируем нормально распределенную совокупность из 660 чисел, стандартным отклонением 1,7 и математическим ожиданием -6,5.

**2.** Сформируем из этой совокупности случайным образом выборку объемом 48 значений (Анализ данных / Выборка).

|  |  |  |  |
| --- | --- | --- | --- |
| 1 | -6,170 | 25 | -5,109 |
| 2 | -3,270 | 26 | -9,878 |
| 3 | -7,342 | 27 | -4,682 |
| 4 | -5,408 | 28 | -6,312 |
| 5 | -7,630 | 29 | -2,637 |
| 6 | -5,522 | 30 | -0,009 |
| 7 | -3,488 | 31 | -4,721 |
| 8 | -6,280 | 32 | -6,366 |
| 9 | -4,092 | 33 | -6,722 |
| 10 | -8,829 | 34 | -3,885 |
| 11 | -7,521 | 35 | -7,180 |
| 12 | -6,216 | 36 | -7,992 |
| 13 | -6,166 | 37 | -5,789 |
| 14 | -7,690 | 38 | -2,964 |
| 15 | -0,516 | 39 | -6,101 |
| 16 | -7,198 | 40 | -5,563 |
| 17 | -6,766 | 41 | -7,564 |
| 18 | -8,521 | 42 | -5,086 |
| 19 | -5,168 | 43 | -6,979 |
| 20 | -7,239 | 44 | -6,179 |
| 21 | -5,633 | 45 | -6,050 |
| 22 | -5,405 | 46 | -8,055 |
| 23 | -6,478 | 47 | -4,882 |
| 24 | -7,555 | 48 | -4,006 |

**3.** Для полученной выборки построить гистограмму частот и график выборочной функции распределения, разбив данные на заданное число интервалов k (двумя способами: провести расчеты на калькуляторе и на компьютере – Анализ данных / Гистограмма).

1) Высчитываем ширину интервала по формуле:

Максимальное значение xmax = -0,009

Минимальное xmin = -9,878

k – количество интервалов = 9;

Ширина интервала:

= 1,257

|  |
| --- |
| Карманы |
| -9,878 |
| -8,621 |
| -7,365 |
| -6,108 |
| -4,852 |
| -3,595 |
| -2,339 |
| -1,082 |
| 0,174 |

Для получения гистограммы частот и графика выборочной функции распределения заходим в Анализ данных / Гистограмма и выбираем нашу выборку с полученными карманами (ставя галочки на вывод графика и интегральный процент)

|  |  |  |
| --- | --- | --- |
| *Карман* | *Частота* | *Интегральный %* |
| -8,78124 | 0,037999 | 4,17% |
| -7,68473 | 0,075998 | 12,50% |
| -6,58821 | 0,208995 | 35,42% |
| -5,49170 | 0,265994 | 64,58% |
| -4,39518 | 0,151997 | 81,25% |
| -3,29867 | 0,075998 | 89,58% |
| -2,20215 | 0,056999 | 95,83% |
| -1,10563 | 0 | 95,83% |
| -0,00892 | 0,037999 | 100,00% |
| Еще | 0 | 100,00% |

0,182

**4.** Найти точечные оценки математического ожидания и дисперсии. Используя эти оценки, нанести кривую нормального распределения на гистограмму частот (двумя способами: на калькуляторе и на компьютере – Анализ данных / Описательная статистика. Кривую распределения на гистограмме построить вручную).

-5,850

Для расчета точечной оценки дисперсии по аналогии используем зависимость:

Откуда:

3,723 3,802

Тогда

1,950

|  |  |
| --- | --- |
| *Столбец1* |  |
|  |  |
| Среднее | -5,85031 |
| Стандартная ошибка | 0,281446 |
| Медиана | -6,16779 |
| Мода | #Н/Д |
| Стандартное отклонение | 1,949913 |
| Дисперсия выборки | 3,80216 |
| Эксцесс | 1,428265 |
| Асимметричность | 0,871482 |
| Интервал | 9,868641 |
| Минимум | -9,87776 |
| Максимум | -0,00912 |
| Сумма | -280,815 |
| Счет | 48 |
| Уровень надежности(95,0%) | 0,566196 |

**5.** Построить 95%-й доверительный интервал для математического ожидания (на калькуляторе с использованием таблиц квантилей).

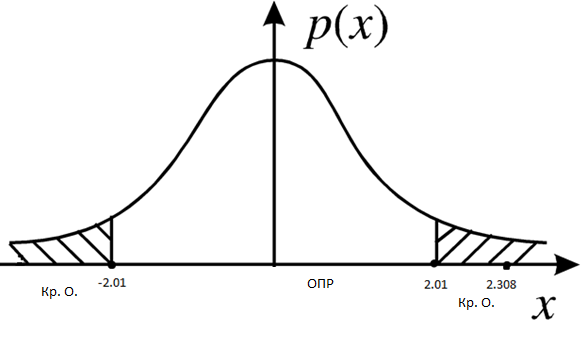
Точечная оценка математического ожидания генеральной совокупности (-5,850) входит в 95% доверительный интервал.

**6.** Проверить гипотезу о том, что математическое ожидание генеральной совокупности равно заданному значению m0 (на калькуляторе).

Проверяемая (нулевая) гипотеза , альтернативная гипотеза . Принимаем по умолчанию вероятность ошибки первого рода α = 0,05.

Положение границы критической области находим двухсторонний критерий

Выборочное значение статистики Стьюдента 2,308 попало в критическую область, нулевая гипотеза отвергается, нельзя считать, что математическое ожидание генеральной совокупности, равно заданному значению



**Рис. 1.1.** Положение критической области

**7.** Разделить выборку на две части (первые n1 значений – первая часть, остальные – вторая). Проверить гипотезу о равенстве средних двумя способами:

1. на калькуляторе, предполагая, что дисперсия генеральной совокупности известна (задано стандартное отклонение σ);
2. на компьютере, в этом случае предполагая, что дисперсия генеральной совокупности неизвестна; вначале проверить гипотезу о равенстве дисперсий (Анализ данных / Двухвыборочный F-тест для дисперсий) и в зависимости от результатов проверки использовать Двухвыборочный t-тест для средних с одинаковыми или различными дисперсиями.

Проверяемая (нулевая) гипотеза , альтернативная гипотеза . Принимаем по умолчанию вероятность ошибки первого рода α = 0,05.

|  |  |  |  |
| --- | --- | --- | --- |
| 1 | -6,16983 | 1 | -5,1682 |
| 2 | -3,27032 | 2 | -7,23947 |
| 3 | -7,34181 | 3 | -5,63278 |
| 4 | -5,40765 | 4 | -5,40479 |
| 5 | -7,6304 | 5 | -6,47787 |
| 6 | -5,52234 | 6 | -7,555 |
| 7 | -3,48838 | 7 | -5,10869 |
| 8 | -6,27992 | 8 | -9,87776 |
| 9 | -4,09187 | 9 | -4,68217 |
| 10 | -8,82911 | 10 | -6,31187 |
| 11 | -7,52114 | 11 | -2,637 |
| 12 | -6,21588 | 12 | -0,00912 |
| 13 | -6,16574 | 13 | -4,7212 |
| 14 | -7,68958 | 14 | -6,36604 |
| 15 | -0,5163 | 15 | -6,7216 |
| 16 | -7,19811 | 16 | -3,8849 |
| 17 | -6,76614 | 17 | -7,18026 |
| 18 | -8,52125 | 18 | -7,99165 |
|  |  | 19 | -5,78896 |
|  |  | 20 | -2,96441 |
|  |  | 21 | -6,10107 |
|  |  | 22 | -5,56304 |
|  |  | 23 | -7,56388 |
|  |  | 24 | -5,08621 |
|  |  | 25 | -6,97894 |
|  |  | 26 | -6,17868 |
|  |  | 27 | -6,0498 |
|  |  | 28 | -8,05518 |
|  |  | 29 | -4,8824 |
|  |  | 30 | -4,0061 |

Рассчитываем статистику:

Положение границ критической области находим для двухстороннего критерия:

Выборочное значение статистики попало в область принятия решения, нулевая гипотеза принимается: можно считать, что средние значения выборок равны.

Проведем расчет при помощи пакета анализ данных:

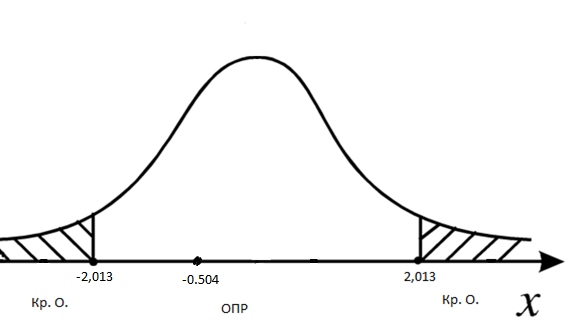
Проверим гипотезу о равенстве дисперсий двух выборок

|  |  |  |
| --- | --- | --- |
|  | *Переменная 1* | *Переменная 2* |
| Среднее | -6,034765 | -5,739634417 |
| Дисперсия | 4,3658847 | 3,569019671 |
| Наблюдения | 18 | 30 |
| df | 17 | 29 |
| F | 1,2232728 |  |
| P(F<=f) одностороннее | 0,3075484 |  |
| F критическое одностороннее | 1,9892843 |  |

Выборочное значение статистики Фишера (1,223) < критического одностороннего (1,989) т.е. попало в область принятия решения следовательно нулевая гипотеза о равенстве дисперсий принимается. Для проверки гипотезы о равенстве средних будем использовать двухвыборочный t-тест с одинаковыми дисперсиями.

|  |  |  |
| --- | --- | --- |
|  | *Переменная 1* | *Переменная 2* |
| Среднее | -6,034765 | -5,739634417 |
| Дисперсия | 4,3658847 | 3,569019671 |
| Наблюдения | 18 | 30 |
| Объединенная дисперсия | 3,8635133 |  |
| Гипотетическая разность средних | 0 |  |
| df | 46 |  |
| t-статистика | -0,5036157 |  |
| P(T<=t) одностороннее | 0,3084674 |  |
| t критическое одностороннее | 1,6786604 |  |
| P(T<=t) двухстороннее | 0,6169349 |  |
| t критическое двухстороннее | 2,0128956 |  |

Выборочное значение статистики Стьюдента (-0,504) попало в область принятия решений, следовательно, гипотеза о равенстве средних принимается: можно считать, что средние значения выборок равны.



**Рис. 1.2.** Положение критической области

**ЗАДАНИЕ 2. ДИСПЕРСИОННЫЙ АНАЛИЗ**

Проверить влияние фактора А на признак Х. В таблице представлены результаты n наблюдений xij (i = 1,2… n) на различных уровнях фактора Аj (j = 1,2… k). Расчет провести двумя способами: на калькуляторе с использованием таблицы квантилей и в пакете Анализ данных / Однофакторный дисперсионный анализ.

|  |  |  |
| --- | --- | --- |
| A1 | A2 | A3 |
| 1,55 | 1,87 | 1,87 |
| 1,45 | 1,75 | 1,56 |
| 1,78 | 1,98 | 1,52 |
| 1,54 | 1,88 | 1,67 |
| 1,7 | 1,8 | 1,59 |

Используем алгоритм однофакторного дисперсионного анализа; имеем количество уровней фактора А (вид технологии) k = 3, количество наблюдений на каждом уровне n = 5.

Групповые средние:

Общее среднее:

Общая сумма квадратов:

Факторная сумма квадратов:

Остаточная сумма квадратов:

Находим по таблице квантиль распределения Фишера:

Выборочное значение статистики Фишера:

Выборочное значение статистики Фишера 6,203 попало в критическую область, нулевая гипотеза о незначимости факторов отклоняется: факторы в рассматриваемой задаче оказывают влияние на точность контролируемого параметра.

Однофакторный дисперсионный анализ:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ИТОГИ |  |  |  |  |  |  |
| *Группы* | *Счет* | *Сумма* | *Среднее* | *Дисперсия* |  |  |
| Столбец 1 | 5 | 8,02 | 1,604 | 0,01773 |  |  |
| Столбец 2 | 5 | 9,28 | 1,856 | 0,00763 |  |  |
| Столбец 3 | 5 | 8,21 | 1,642 | 0,01927 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| Дисперсионный анализ |  |  |  |  |  |  |
| *Источник вариации* | *SS* | *df* | *MS* | *F* | *P-Значение* | *F критическое* |
| Между группами | 0,184573333 | 2 | 0,092286667 | 6,203450594 | 0,014125753 | 3,885293835 |
| Внутри групп | 0,17852 | 12 | 0,014876667 |  |  |  |
|  | 0 |  |  |  |  |  |
| Итого | 0,363093333 | 14 |  |  |  |  |

**ЗАДАНИЕ 3. ПАРНАЯ РЕГРЕССИЯ**

Для заданной выборки:

1. Построить диаграмму рассеяния.
2. Используя калькулятор, найти параметры парной линейной регрессионной модели.
3. Нанести линию регрессии на диаграмму рассеяния.
4. Проверить значимость построенной модели.
5. Вычислить коэффициент детерминации.
6. Используя инструмент Линия тренда Excel, опробовать построение линейной и других вариантов регрессионных моделей: экспоненциальной, степенной, логарифмической, полиномиальных 2–4 степеней. Результат для каждой модели представить графически: диаграмма рассеяния, линия регрессии, ее уравнение, коэффициент детерминации. Выбрать регрессию, наиболее адекватную опытным данным.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| xi | 4,1 | 4,3 | 4,5 | 4,7 | 4,9 | 5,1 | 5,3 | 5,5 |
| yi | 7,47 | 7,22 | 7,04 | 6,89 | 6,85 | 6,47 | 6,15 | 5,91 |

Объем выборки n = 8. Выборочные средние

Найдем оценки параметров линейной регрессии:

Тогда:

Уравнение линейной регрессии:

Выполним проверку, сравнив суммы yi и прогнозируемых значений yi

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ~yi | 7,501 | 7,286 | 7,072 | 6,857 | 6,643 | 6,428 | 6,214 | 5,999 |

Диаграмма рассеяния и расчетная прямая показаны на рис. 3.1

Рис. 3.1. Диаграмма рассеяния и линия регрессии

Проверим значимость регрессии:

Критическое значение статистики Фишера:

Гипотеза о незначимости отклоняется, регрессионная модель значима.

Найдем соответствующий коэффициент детерминации:

Другие примеры Диаграмм рассеяния:

Сравнивая значения коэффициентов детерминации, видим, что полиномиальная модель четвертой степени существенно лучше отображает результаты наблюдений. Это же видно и из графиков.

**ЗАДАНИЕ 4. МНОЖЕСТВЕННАЯ РЕГРЕССИЯ**

Найти параметры множественной линейной регрессии. Проверить значимость полученной модели и значимость каждого фактора. Найти коэффициент детерминации. Расчет провести, используя инструмент Регрессия пакета Анализ данных Excel.

|  |  |  |  |
| --- | --- | --- | --- |
| x1 | x2 | x3 | Y |
| 1 | 0,8 | 9 | 120 |
| 3 | 1 | 12 | 190 |
| 5 | 1,4 | 21 | 310 |
| 7 | 1,9 | 13 | 290 |
| 9 | 2,4 | 21 | 400 |
| 11 | 2,2 | 24 | 450 |
| 13 | 2,7 | 35 | 600 |
| 15 | 2,5 | 28 | 560 |
| 17 | 3 | 15 | 490 |
| 19 | 3,2 | 35 | 710 |
| 21 | 2,6 | 28 | 660 |
| 23 | 2,4 | 35 | 720 |
| 25 | 2 | 15 | 600 |
| 27 | 1,5 | 24 | 700 |
| 29 | 3 | 25 | 770 |

Модель имеет вид:

Значимость F (1,0176E-15) << 0,05, следовательно, модель значима. Также все P значения для переменных << 0,05, следовательно, все факторы также значимы

Коэффициент детерминации R^2 = 0,9984.