МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное АВТОНОМНОЕ образовательное учреждение высшего профессионального образования
«Национальный исследовательский ядерный университет «МИФИ»

**Димитровградский инженерно-технологический институт –**филиал федерального государственного автономного образовательного учреждения высшего
профессионального образования «Национальный исследовательский ядерный университет «МИФИ»
**(ДИТИ НИЯУ МИФИ)**

**Лабораторная работа №2
по курсу «Основы математической статистики и
планирование эксперимента»**

Составил: доцент кафедры
 высшей математики
 канд. экон. наук
 Кожухова В.Н.

Димитровград 2017

**Лабораторная работа №2
по курсу «Основы математической статистики и
планирование эксперимента»**

**Исходные данные:**

Рассматриваются две нормально распределенные случайные величины (СВ) *X* и *Y*, для которых заданы истинные математическое ожидание (МО) *M*, среднеквадратическое отклонение (СКО) .

Для каждой СВ имеется выборка объемом *n* = 30 наблюдений:

,

.

**Задание:**

1. Рассчитать выборочное среднее *m*, стандартное отклонение *S*, Сравнить их с истинными характеристиками СВ.
2. Построить доверительные интервалы для оценок математического ожидания и дисперсии при заданном уровне значимости , считая истинные значения математического ожидания и дисперсии а) известными б) неизвестными. Проверить гипотезу о попадании истинного значения в построенный доверительный интервал.
3. Построить гистограмму распределения и график теоретической плотности распределения (на основе выборочных характеристик), эмпирическую и теоретическую функции распределения для *X* и *Y*. Наложить их друг на друга.
4. Проверить равенство математических ожиданий двух нормально распределенных случайных величин X и Z на основе имеющихся выборок, считая, что дисперсии а) равны и известны; б) равны и неизвестны.
5. Проверить равенство дисперсий двух нормально распределенных случайных величин X и Z на основе имеющихся выборок.

**Обозначения в лабораторной работе:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Теоретическое (истинное) значение | Выборочные значения, рассчитанные без группировки | Выборочные значения, рассчитанные на основе группированных данных | Выборочная дисперсия при известном генеральном среднем |
| X | M(X), σ2(X), σ(X) | , ,  | , ,  |  |

Аналогично для Y и Z.

**Ход выполнения лабораторной работы:**

1. Рассчитать выборочное среднее, выборочную дисперсию и стандартное отклонение по формулам для выборок из распределения случайных величин X и Y:

 - выборочное среднее,

 - несмещенная выборочная дисперсия,

 - несмещенное выборочное СКО,

где  – ряд наблюдений в выборке из распределения случайной величины X.

Аналогично для Y:

 - выборочное среднее,

 - несмещенная выборочная дисперсия,

 - несмещенное выборочное СКО,

где  – ряд наблюдений в выборке из распределения случайной величины Y.

2. Расчет доверительного интервала проделать и для *X*, и для *Y*.

2.1 Рассмотрим задачу определения доверительного интервала для математического ожидания (генеральной средней) нормальной выборки а) при известной генеральной дисперсии; б) при неизвестной генеральной дисперсии.

а) Задаваясь значением двусторонней доверительной вероятности γ, определим верхнюю и нижнюю доверительные границы:

,

где  – аргумент функции Лапласа , *п* – объем выборки;  – оценка математического ожидания, определенная по выборке.

**Найденный таким образом интервал с вероятностью γ = 1 – α накроет неизвестную генеральную среднюю.**

|  |  |
| --- | --- |
|  | Для расчета  можно воспользоваться встроенной функцией Excel. Известно, что функция Лапласа , где  – интегральная функция стандартного нормального распределения (с нулевым математическим ожиданием и единичной дисперсией), тогда  = НОРМ.ОБР(γ;0;1). |

б) При неизвестной генеральной дисперсии:

,

где  – квантили распределения Стьюдента с *k* = *n*–1 степенями свободы, *п* – объем выборки; *, S* – оценки математического ожидания и среднего квадратического отклонения, определенные по выборке.

**Найденный таким образом интервал с вероятностью γ накроет неизвестную генеральную среднюю.**

|  |  |
| --- | --- |
|  | Для расчета  можно воспользоваться встроенной функцией Excel ==СТЬЮДЕНТ.ОБР (порядок квантиля; n-1). |

2.2 Определение доверительного интервала для дисперсии нормальной выборки а) при известном генеральном среднем; б) при неизвестном генеральном среднем.

**а) При известном генеральном среднем найдем выборочную дисперсию следующим образом:**

,

где  – известное теоретическое значение математического ожидания X.

Тогда доверительный интервал для дисперсии:



где  – квантили распределения хи-квадрат с n степенями свободы.

|  |  |
| --- | --- |
|  | Можно воспользоваться встроенной функцией Excel  = ХИ2.ОБР  и = ХИ2.ОБР. |

**б) при неизвестном генеральном среднем:**

.

Разница лишь в том, что мы берем оценку дисперсии, полученную в первом пункте.

**Внимание: п.2 проделать и для Y тоже!**

3. Построить гистограмму распределения и график теоретической плотности распределения (на основе выборочных характеристик), эмпирическую и теоретическую функции распределения **для *X,* идля *Y***. Наложить их друг на друга.

а) Построить вариационный ряд – упорядочить значения в выборке по возрастанию.

б) Разбить полученный ряд на *m* интервалов равной ширины *w* и подсчитать количество значений, попавших в каждый интервал (эмпирические абсолютные частоты).

Минимальное число столбцов (интервалов) гистограммы распределения (формула Стерджесса): .

Ширина интервала .

|  |  |
| --- | --- |
|  | Минимальное и максимальное значения в выборке: МИН(число1;число2; ...) и МАКС(число1;число2; ...)Количество наблюдений, попавших в *i*-тый интервал, определяется с помощью функции ЧАСТОТА(массив\_данных;массив\_интервалов). Массив\_интервалов – правые границы каждого интервала! Формулу необходимо ввести как формулу массива. После ввода формулы в первую ячейку (например, А1) выделите весь нужный диапазон (например, A1:A6 – для 6 интервалов), нажмите клавишу F2, а затем нажмите клавиши CTRL+SHIFT+ENTER. Если формула не будет введена как формула массива, отобразится только одно ее значение в ячейке A1.  |

в) Рассчитать высоту столбцов и построить гистограмму эмпирического распределения

Высота *i*-того столбца: , где  – количество наблюдений, попавших в *i*-тый интервал (столбец частоты), *h* – ширина интервала, *n* – число наблюдений в выборке.

г) Рассчитать накопленные относительные частоты попадания в интервалы и построить интегральную функцию эмпирического распределения (э.ф.р.).

Гистограмму и функцию распределения необходимо строить по серединам интервалов (подписи по оси X). Эмпирическая функция распределения представит собой диапазон накопленных частот .

**В п. 3 нужно рассчитать эмпирические числовые характеристики по группированным данным:**

 – выборочное среднее по группированным данным,  – середины интервалов;

 – несмещенная выборочная дисперсия по группированным данным,

 – несмещенное выборочное СКО по группированным данным.

|  |  |
| --- | --- |
|  | Воспользуйтесь формулой =СУММПРОИЗВ(массив ; массив  ) /(n –1). Массив  нужно рассчитать как отдельный столбец. |

д) Рассчитать значения **теоретической** плотности вероятности для нормально распределенной СВ:

. (\*)

где принять . Аргументом плотности будут середины интервалов.

|  |  |
| --- | --- |
|  | (\*) Можно воспользоваться функцией =**НОРМ.РАСП** (;;;0)**.** |

Пусть  – значения теоретической плотности вероятностей, рассчитанные по формуле (\*). Тогда теоретические частоты нормального распределения .



Рисунок 1 – Гистограмма распределения и плотность распределения

**Эмпирическая функция распределения (кумулята) представляет собой диапазон накопленных частот .**

Формула теоретической функции нормального распределения:

, (\*\*)

где принять .

|  |  |
| --- | --- |
|  | (\*\*) Можно воспользоваться функцией =**НОРМ.РАСП** (;;;1)**.** |

Аргумент интегральной функции нормального распределения – середины интервалов.



Рисунок 2 – Эмпирическая и теоретическая функции распределения

4. Проверить равенство математических ожиданий двух нормально распределенных случайных величин X и Z, считая, что дисперсии а) равны и известны б) равны и неизвестны.

В п. 4 и 5 рассматривается двусторонняя критическая область.

Проверим гипотезу  при альтернативной гипотезе .

**а) Пусть дисперсии *X* и *Z* равные и известны: .**

Рассмотрим случайную величину , которая будет статистикой критерия. Здесь *m* – объем выборки для *Z.*

Если , то гипотеза о равенстве математических ожиданий отвергается. Здесь  – квантиль порядка (1 – α/2) стандартного нормального распределения. Стандартное нормальное распределение симметрично относительно нуля, поэтому нужно находить только одну критическую точку. Гипотеза принимается, если .

|  |  |
| --- | --- |
|  | =НОРМ.СТ.ОБР(1 – α/2) |

**б) Пусть дисперсии *X* и *Z* равны ** и неизвестны.**

.

Рассмотрим случайную величину , которая будет статистикой критерия.

Если , то гипотеза о равенстве математических ожиданий отвергается. Здесь  – квантиль порядка (1 – α/2) распределения Стьюдента с  степенями свободы. Распределение Стьюдента симметрично относительно нуля, поэтому нужно находить только одну критическую точку. Гипотеза принимается, если .

|  |  |
| --- | --- |
|  | =СТЬЮДЕНТ.ОБР(1 – α/2; n + m – 2) |

5. Проверить равенство дисперсий двух нормально распределенных случайных величин *X* и *Z*.

Проверим гипотезу  при альтернативной гипотезе .

Рассмотрим случайную величину , которая будет статистикой критерия Фишера. Здесь  – большая дисперсия,  – меньшая дисперсия из выборочных дисперсий *X* и *Z*.

Если  или , то гипотеза о равенстве дисперсий отвергается. Здесь  – квантили порядка (α/2) и (1 – α/2) распределения Фишера с *k*1 и *k*2 степенями свободы. Распределение Фишера асимметрично, поэтому нужны две критических точки.

Гипотеза принимается, если .

k1 = объем выборки с бОльшей дисперсией минус 1; k2 = объем выборки с меньшей дисперсией минус 1.

|  |  |
| --- | --- |
|  | =F.ОБР(α/2; k1; k2); =F.ОБР(1 – α/2; k1; k2) |