

Контрольная работа по эконометрике Вариант 8

Задание 2.

1. Представьте информацию о переменной, используя надстройку «Анализ данных» Описательные статистики. *Что за характеристика может быть представлена такими данными.* Постройте гистограмму с использованием диаграмм Excel.
2. Постройте функцию плотности распределения. Нанесите линию на гистограмму. (2 балл) На основании визуального анализа, можно ли предполагать, что случайная величина распределена по нормальному закону распределения.
3. Проверьте ряд данных на наличие аномальных значений. Если аномальные значения идентифицированы, исключите их из выборки. Повторите п.1-4 после исключения аномальных значений. *Сильно ли изменились результаты ?* (2 балл)
4. Проверьте ряд данных, полученный после исключения аномалий, на нормальный закон распределения с использованием коэффициентов асимметрии и эксцесса. *Подтвердился ли результат визуального анализа.* (2 балла)

Решение:

1. Воспользуемся функцией Описательная статистика пакета Анализ данных MS Excel:

Столбец1	
Среднее	-0,05857
Стандартная ошибка	0,014893
Медиана	0,025597
Мода	0,556451
Стандартное отклонение	1,290304
Дисперсия выборки	1,664884
Эксцесс	145,3864
Асимметричность	-9,73136
Интервал	27,51016
Минимум	-23,6325
Максимум	3,877662
Сумма	-439,595
Счет	7506

С помощью описательной статистики представлены показатели центра распределения и вариации. Ряд распределения содержит 7506 значений. Их сумма равна -439,595; среднее значение -0,05857; типичное 0,556451; срединное 0,025597.

Для группировки используем самостоятельный расчет интервалов по формулам ниже.

Количество интервалов (n) определяется по формуле Стерджесса:

$$n = 1 + 3,322 \lg N, \text{ где}$$

N – объём совокупности.

$$n = 1 + 3,322 \cdot \lg(7506) = 13,87 \approx 14$$

Поможем с заданиями по эконометрике: https://www.matburo.ru/sub_subject.php?p=eckr

Величина варианты (интервала) зависит от размаха вариации и количества групп, на которые разбивается совокупность. Значение варианты рассчитывается по формуле:

$$i = \frac{x_{\max} - x_{\min}}{n}, \text{ где}$$

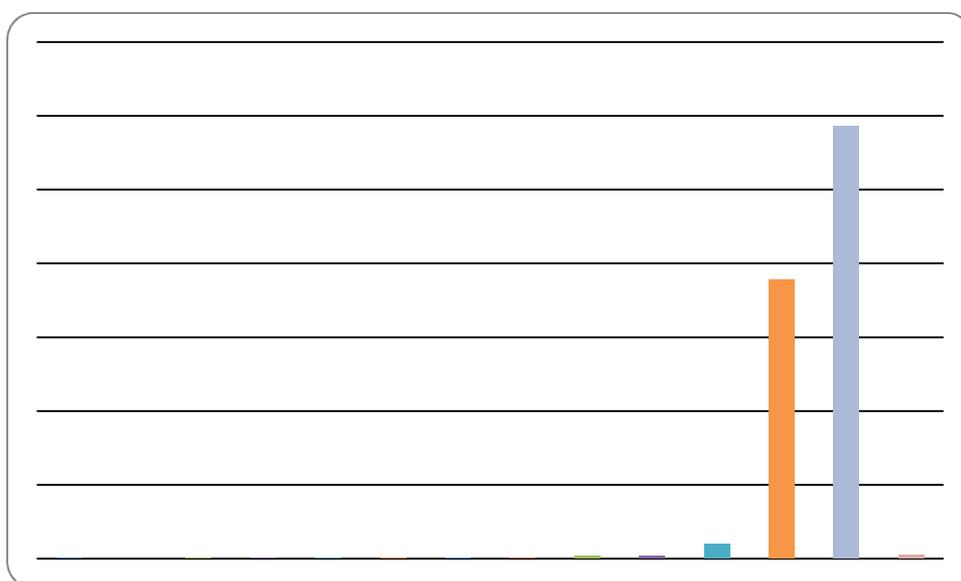
x_{\max} , x_{\min} – соответственно максимальное и минимальное значение факторного признака.

$$i = \frac{3,877662 - (-23,6325)}{14} = 1,965012$$

Построим группировку:

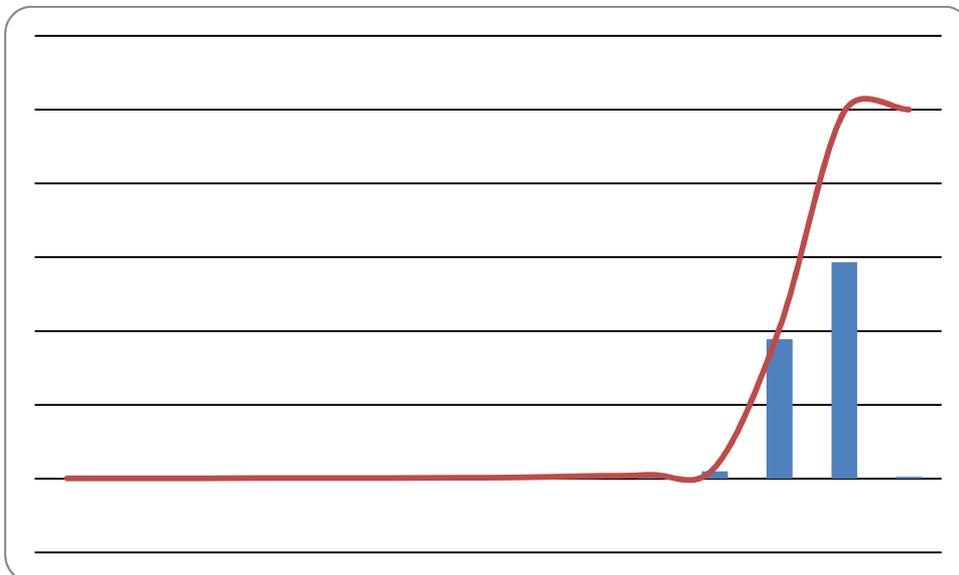
Левая граница	Правая граница	Частота	Относительная частота	Функция плотности распределения
-23,6325024	-21,66749069	9	0,001199041	0,001199041
-21,66749069	-19,70247897	0	0	0,001199041
-19,70247897	-17,73746726	1	0,000133227	0,001332268
-17,73746726	-15,77245555	1	0,000133227	0,001465494
-15,77245555	-13,80744384	1	0,000133227	0,001598721
-13,80744384	-11,84243212	4	0,000532907	0,002131628
-11,84243212	-9,87742041	4	0,000532907	0,002664535
-9,87742041	-7,912408697	8	0,001065814	0,003730349
-7,912408697	-5,947396984	25	0,003330669	0,007061018
-5,947396984	-3,982385271	28	0,003730349	0,010791367
-3,982385271	-2,017373558	148	0,019717559	0,030508926
-2,017373558	-0,052361846	2837	0,377964295	0,408473221
-0,052361846	1,912649867	4401	0,586330935	0,994804157
1,912649867	3,87766158	39	0,005195843	1
Итого:		7506	1	

Построим гистограмму:



Поможем с заданиями по эконометрике: https://www.matburo.ru/sub_subject.php?p=eckr

2. Добавим функцию плотности распределения:



Вывод: по графикам можно сделать предположение, что распределение случайной величины не является нормальным.

3. Для того, чтобы проверить имеются ли резко выделяющиеся наблюдения или же убрать резко выделяющиеся наблюдения (при $V > 33,3\%$), необходимо использовать правило «трех сигма», которое заключается в соблюдении следующего интервала:

$$x_i \in [\bar{x} - 3\sigma_x; \bar{x} + 3\sigma_x]$$

$$x_i \in [-0,05857 - 3 \cdot 1,290304; -0,05857 + 3 \cdot 1,290304]$$

$$x_i \in [-3,929477; 3,812346]$$

Упорядочим данные по возрастанию и исключим аномальные значения. Построим описательную статистику:

Столбец1	
Среднее	0,03964
Стандартная ошибка	0,007586
Медиана	0,029425
Мода	0,556451
Стандартное отклонение	0,653612
Дисперсия выборки	0,427208
Эксцесс	7,294483
Асимметричность	-1,28492
Интервал	7,302389
Минимум	-3,85816
Максимум	3,444228
Сумма	294,2475
Счет	7423

Поможем с заданиями по эконометрике: https://www.matburo.ru/sub_subject.php?p=eckr

Мода и медиана почти не изменились, у среднего значения изменился знак. Заметно уменьшились стандартное отклонение, дисперсия выборки, асимметрия и эксцесс.

Выполним расчёты для построения новой гистограммы:

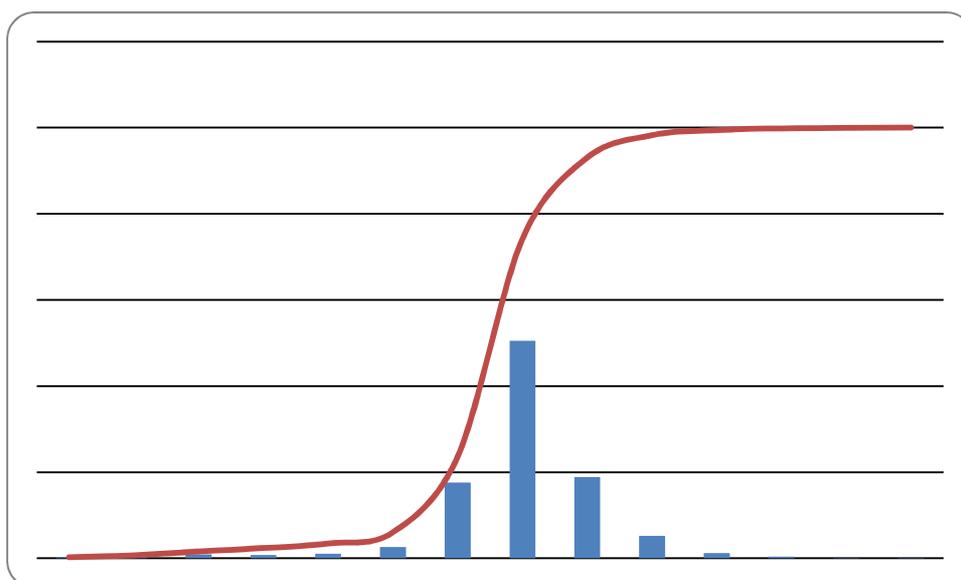
$$n = 1 + 3,322 \cdot \lg(7423) = 13,86 \approx 14$$

$$i = \frac{3,444228 - (-3,85816)}{14} = 0,521599$$

Построим группировку:

Левая граница	Правая граница	Частота	Относительная частота	Функция плотности распределения
-3,85816046	-3,33656128	19	0,002559612	0,002559612
-3,33656128	-2,81496209	35	0,004715075	0,007274687
-2,81496209	-2,2933629	66	0,008891284	0,016165971
-2,2933629	-1,77176372	56	0,00754412	0,02371009
-1,77176372	-1,25016453	79	0,010642597	0,034352688
-1,25016453	-0,72856534	194	0,026134986	0,060487673
-0,72856534	-0,20696616	1307	0,176074363	0,236562037
-0,20696616	0,31463303	3751	0,505321299	0,741883336
0,31463303	0,836232217	1400	0,188602991	0,930486326
0,836232217	1,357831404	387	0,052135255	0,982621582
1,357831404	1,879430591	89	0,011989762	0,994611343
1,879430591	2,401029778	28	0,00377206	0,998383403
2,401029778	2,922628964	7	0,000943015	0,999326418
2,922628964	3,444228151	5	0,000673582	1
Итого:		7423	1	

Построим гистограмму и функцию плотности распределения:



В данном случае визуальный анализ позволяет сделать вывод о нормальном распределении.

Поможем с заданиями по эконометрике: https://www.matburo.ru/sub_subject.php?p=eckr

4. Оценка существенности асимметрии проводится с помощью среднеквадратической ошибки:

$$\sigma_{As} = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}} = \sqrt{\frac{6(7423-1)}{(7423+1)(7423+3)}} = 0,028421$$

Так как отношение коэффициента асимметрии по модулю к средней квадратической ошибке меньше 3 ($1,28492/0,028421 = 45,21 > 3$), то асимметрия существенна.

Средняя квадратическая ошибка эксцесса определяется по формуле:

$$\sigma_{Es} = \sqrt{\frac{24n(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}} = \sqrt{\frac{24*7423*(7423-2)(7423-3)}{(7423-1)^2(7423+3)(7423+5)}} = 0,056819$$

Так как отношение коэффициента эксцесса по модулю к средней квадратической ошибке меньше 3 ($7,294483/0,056819 = 128,38 > 3$), то эксцесс существенен.

Следовательно, закон распределения не является нормальным.

Задание 3.

1. Разделите ряд данных, полученный после исключения аномалий, на 2 подвыборки примерно равного объема. Проверьте гипотезу о равенстве средних и дисперсий двух сформированных подвыборок. Выполните проверку для различных значений доверительной вероятности (0,95;0,9;0,99). *Изменился ли вывод при разных значениях доверительной вероятности.* (2 балла)
2. По ряду данных, полученных после исключения аномалий, найдите точечную и интервальную оценки среднего. Найдите доверительные интервалы для различных значений доверительной вероятности (0,95;0,9;0,99). *При каком значении доверительной вероятности доверительный интервал шире. Почему?* (2 балл)

Решение:

1. Разделим ряд данных на две подвыборки со следующими характеристиками:

n_1	n_2	\bar{x}	\bar{y}	σ_x^2	σ_y^2
3711	3712	0,028288	0,050988	0,431371	0,422674

Проверим гипотезу $H_0: \bar{x}_0 = \bar{y}_0$. В качестве альтернативной (конкурирующей) гипотезы можно взять $H_1: \bar{x}_0 \neq \bar{y}_0$.

Так как проверяется гипотеза о равенстве средних, то в качестве критерия проверки нулевой гипотезы следует принять случайную величину

$$z_{набл} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}}} = \frac{0,028288 - 0,050988}{\sqrt{\frac{0,431371}{3711} + \frac{0,422674}{3712}}} = -1,49645$$

По таблице функции Лапласа найдем критическую точку при доверительной вероятности 0,95 из условия:

Поможем с заданиями по эконометрике: https://www.matburo.ru/sub_subject.php?p=eckr

$$\Phi(z_{кр}) = \frac{1-\alpha}{2} = \frac{1-0,05}{2} = 0,475, \text{ откуда } z_{кр} = 1,96.$$

Так как $|z_{набл}| < z_{кр}$, то нулевую гипотезу $H_0: \bar{x}_0 = \bar{y}_0$ следует принять.

Проверим гипотезу при доверительной вероятности 0,99:

$$\Phi(z_{кр}) = \frac{1-\alpha}{2} = \frac{1-0,01}{2} = 0,495, \text{ откуда } z_{кр} = 2,58.$$

Так как $|z_{набл}| < z_{кр}$, то нулевую гипотезу $H_0: \bar{x}_0 = \bar{y}_0$ следует принять. Вывод не изменился.

Проверим гипотезу о равенстве дисперсий. Используем критерий Фишера-Снедекора. Найдем отношение большей дисперсии к меньшей:

$$F_{набл} = \frac{0,431371}{0,422674} = 1,020575$$

При уровне значимости $\alpha = 0,05$ и числам степеней свободы $k_1 = n_1 - 1 = 3711 - 1 = 3710$ (большая дисперсия) и $k_2 = n_2 - 1 = 3712 - 1 = 3711$ (меньшая дисперсия) найдем критическую точку $F_{кр}(0,05;3710;3711) = 1,055498$. Так как $F_{набл} < F_{кр}$, то гипотезу о равенстве дисперсий следует принять.

При уровне значимости 0,99 $F_{кр}(0,01;3710;3711) = 1,07939$. Так как $F_{набл} < F_{кр}$, то гипотезу о равенстве дисперсий следует принять. Вывод не изменился.

2. Точечная оценка среднего составляет 0,03964.

В предположении нормальности распределения интервальная оценка среднего задаётся формулой:

$$P = \left\{ \bar{x} - t_{1-\alpha;n-1} \frac{\sigma}{\sqrt{n}} < MX < \bar{x} + t_{1-\alpha;n-1} \frac{\sigma}{\sqrt{n}} \right\} = \alpha$$

При $\alpha = 0,95$, $n = 7423$ критическая точка $t_{1-\alpha;n-1} = 1,96$ и поскольку $\bar{x} = 0,03964$, $\sigma_x = 0,653612$, 95%-ная интервальная оценка MX принимает вид:

$$0,03964 - 1,96 \frac{0,653612}{\sqrt{7423}} < MX < 0,03964 + 1,96 \frac{0,653612}{\sqrt{7423}} \text{ или } 0,024769 < MX < 0,054511$$

При $\alpha = 0,99$ критическая точка $t_{1-\alpha;n-1} = 2,58$, 99%-ная интервальная оценка MX принимает вид:

$$0,03964 - 2,58 \frac{0,653612}{\sqrt{7423}} < MX < 0,03964 + 2,58 \frac{0,653612}{\sqrt{7423}} \text{ или } 0,020094 < MX < 0,059186$$

При доверительной вероятности 0,99 интервал шире. Это связано со снижением точности оценки.