**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ**

**РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ**

**УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ**

**«МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ**

(национальный исследовательский университет)»



**ОТЧЕТ**

**по лабораторным работам №1-3**

по дисциплине:

«Теория вероятностей и статистика в машинном обучении»

| Исполнитель студент группы М8О-110М-23 |  | Николаев В.А. |
| --- | --- | --- |

Москва, 2023

**СОДЕРЖАНИЕ**

[Лабораторная работа №1 3](#_30j0zll)

[Часть 1 3](#_1fob9te)

[Часть 2 8](#_3znysh7)

[Лабораторная работа №2 15](#_2et92p0)

[Лабораторная работа №3 22](#_tyjcwt)

[Часть 1 22](#_3dy6vkm)

[Часть 2 23](#_1t3h5sf)

[Список использованных источников 31](#_4d34og8)

# ЛАБОРАТОРНАЯ РАБОТА №1

## Часть 1

*1. Найти на сайте РосСтат данные, распределение которых было бы близко к*

*a) нормальному (визуально выглядит как кривая Гаусса - «колокол»),*

Выбранные данные для анализа - распределение населения по величине среднедушевых денежных доходов за 2020 г. – представлены в Таблице 1[[1]](#footnote-0).

Таблица 1. Распределение населения по величине среднедушевых денежных доходов за 2020 г.

| Уровень дохода | Доля населения, % |
| --- | --- |
| до 7 000,0 | 3,5 |
| от 7 000,1 до 10 000,0 | 5,6 |
| 10 000,1-14 000,0 | 9,6 |
| 14 000,1-19 000,0 | 12,8 |
| 19 000,1-27 000,0 | 17,9 |
| 27 000,1-45 000,0 | 21,2 |
| 45 000,1-60 000,0 | 14,7 |
| 60 000,1-75 000,0 | 5,8 |
| 75 000,1-100 000,0 | 4,7 |
| свыше 100 000,0 | 4,2 |

*б) равномерному*

Выбранные данные для анализа - средняя заработная плата по 10-процентным группам работников организаций (без субъектов малого предпринимательства) за 2022 г. – представлены в Таблице 2[[2]](#footnote-1).

Таблица 2. Средняя заработная плата по 10-процентным группам работников организаций (без субъектов малого предпринимательства) за 2022 г.

| Средний уровень з/п, руб. | Доля, % |
| --- | --- |
| 18 146 | 10% |
| 27 165 | 10% |
| 34 323 | 10% |
| 41 201 | 10% |
| 48 497 | 10% |
| 57 008 | 10% |
| 67 962 | 10% |
| 83 131 | 10% |
| 109 194 | 10% |
| 238 278 | 10% |

*2. Найти распределение со смещенной медианой относительно среднего (~15% размаха) и несмещенной.*

Выбранные данные, имеющие распределение со смещенной медианой относительно среднего - распределение браков по возрасту невесты за 2004 г. – представлены в Таблице 3[[3]](#footnote-2).

Таблица 3. Распределение браков по возрасту невесты за 2004 г.

| Возраст невесты | Кол-во |
| --- | --- |
| до 18 | 23 428 |
| 18-24 | 519 606 |
| 25-34 | 276 317 |
| 35 и более | 160 167 |

Данные с распределением с несмещенной медианой относительно среднего совпадают с нормальным распределением и данными Таблицы 1.

*3. Посчитать описательные характеристики распределения выбранных данных (среднее, мода, медиана), дать визуальное представление данных (условное форматирование Excel, построить график рассеивания/гистограмму/круговую диаграмму).*

Визуальное представление выборки №1 представлено на Рисунке 1.

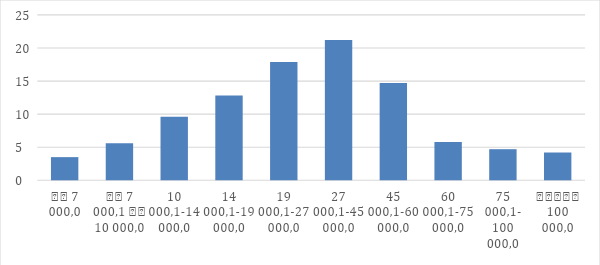


Рисунок 1. Распределение населения по величине среднедушевых денежных доходов за 2020 г.

Вспомогательные данные для расчета показателей представлены в Таблице 4.

Таблица 4. Вспомогательные данные для расчета показателей выборки №1

| Группы | Середина интервала, xцентр | Кол-во, fi | xi\*fi | Накопленная частота, S |
| --- | --- | --- | --- | --- |
| 0 - 7000 | 3500 | 4 | 12250 | 4 |
| 7000,1 - 10000 | 8500,05 | 6 | 47600,28 | 9 |
| 10000,1 - 14000 | 12000,05 | 10 | 115200,48 | 19 |
| 14000,1 - 19000 | 16500,05 | 13 | 211200,64 | 32 |
| 19000,1 - 27000 | 23000,05 | 18 | 411700,895 | 49 |
| 27000,1 - 45000 | 36000,05 | 21 | 763201,06 | 71 |
| 45000,1 - 60000 | 52500,05 | 15 | 771750,735 | 85 |
| 60000,1 - 75000 | 67500,05 | 6 | 391500,29 | 91 |
| 75000,1 - 100000 | 87500,05 | 5 | 411250,235 | 96 |
| 100000,1 - 150000 | 125000,05 | 4 | 525000,21 | 100 |
| Итого |  | 100 | 3 660 654,825 |  |

Средняя взвешенная:

*=*==36 606,55 (руб.)

Мода:

=, где – начало модального интервала; h – величина интервала; –частота, соответствующая модальному интервалу; – предмодальная частота; – послемодальная частота.

=27 000,1+17999,9\*=33061,29 (руб.)

Медиана:

==27 000,1+\*(=27 509,53 (руб.)

Визуальное представление выборки №2 представлено на Рисунке 2.

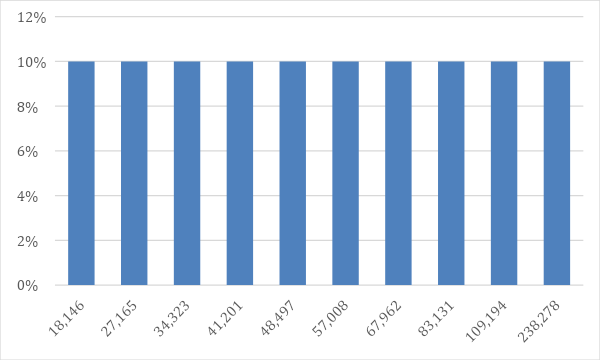


Рисунок 2. Средняя заработная плата по 10-процентным группам работников организаций (без субъектов малого предпринимательства) за 2022 г.

Средняя взвешенная:

*===* 72 491 (руб.)

Мода отсутствует (имеются несколько показателей с одинаковым значением).

Для расчета медианы находим , при котором накопленная частота S будет больше = 50. Это значение = 57008. Таким образом, медиана равна 57 008 руб.

Визуальное представление выборки №3 представлено на Рисунке 3.

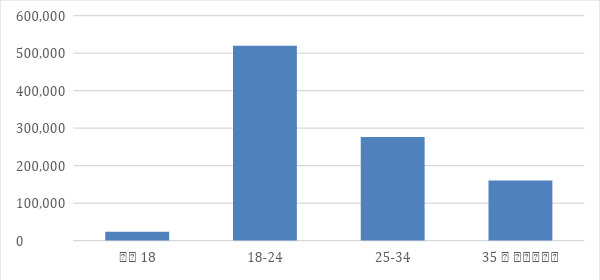


Рисунок 3. Распределение браков по возрасту невесты за 2004 г.

Вспомогательные данные для расчета показателей представлены в Таблице 5.

Таблица 5. Вспомогательные данные для расчета показателей выборки №2

| Группы | Середина интервала, xцентр | Кол-во, fi | xi\*fi | Накопленная частота, S |
| --- | --- | --- | --- | --- |
| 0 - 18 | 9 | 23 428 | 210 852 | 23 428 |
| 18 - 24 | 21 | 519 606 | 10 911 726 | 543 034 |
| 25 - 34 | 29.5 | 276 317 | 8 151 351,5 | 819 351 |
| 35 - 50 | 42.5 | 160 167 | 6 807 097,5 | 979 518 |
| Итого |  | 979 518 | 26 081 027 |  |

Средняя взвешенная:

*===*26,63 (лет)

Мода:

=, где – начало модального интервала; h – величина интервала; –частота, соответствующая модальному интервалу; – предмодальная частота; – послемодальная частота.

=18+6\*=22,03 (лет)

Медиана:

==18+\*(=23,39 (лет)

## Часть 2

*Найти датасет с аномальными значениями, которые можно выявить, построив диаграмму размаха «ящик с усами». Попробовать выдвинуть гипотезу о причинах аномальности, которую можно подтвердить дальнейшими исследованиями, используя синтетические признаки (или корреляционный анализ и т.п.).*

Выбранный показатель – средняя цена 1 кв.м. в разбивке по районам Москвы за ноябрь 2023 г[[4]](#footnote-3). Наблюдения представлены в Таблице 6.

Таблица 6. Данные по средней цене 1 кв.м. в районах Москвы

| № | Район | Цена 1 кв. м. |
| --- | --- | --- |
| 1 | Остоженка | 529 798 |
| 2 | Арбат | 491 032 |
| 3 | Якиманка | 488 403 |
| 4 | Тверской | 471 041 |
| 5 | Центр Москвы | 468 699 |
| 6 | Хамовники | 436 960 |
| 7 | Пресненский | 391 878 |
| 8 | Мещанский | 391 147 |
| 9 | Замоскворечье | 390 349 |
| 10 | Красносельский | 381 260 |
| 11 | Беговой | 375 676 |
| 12 | Таганский | 374 585 |
| 13 | Дорогомилово | 366 909 |
| 14 | Гагаринский, Ломоносовский, Раменки | 353 391 |
| 15 | Донской | 351 960 |
| 16 | Хорошевский | 325 261 |
| 17 | Аэропорт, Сокол | 320 253 |
| 18 | Алексеевский | 315 348 |
| 19 | Марьина роща, Савеловский | 314 779 |
| 20 | Филевский парк | 314 468 |
| 21 | Проспект Вернадского | 314 197 |
| 22 | Басманный | 313 096 |
| 23 | Академический | 307 157 |
| 24 | Динамо | 306 647 |
| 25 | Сокольники | 306 595 |
| 26 | Черемушки | 304 532 |
| 27 | Хорошево-Мневники | 304 042 |
| 28 | Южнопортовый | 302 435 |
| 29 | Нижегородский | 296 601 |
| 30 | Крылатское | 289 663 |
| 31 | Тропарево-Никулино | 278 883 |
| 32 | Свиблово | 278 678 |
| 33 | Можайский, Фили-Давыдково | 278 512 |
| 34 | Даниловский, Котловка | 277 068 |
| 35 | Кунцево | 277 056 |
| 36 | Останкинский, Ростокино | 276 486 |
| 37 | Покровское-Стрешнево, Щукино | 273 189 |
| 38 | Очаково-Матвеевское | 272 741 |
| 39 | Соколиная гора | 272 303 |
| 40 | Нагатино-Садовники, Нагатинский Затон | 271 690 |
| 41 | Зюзино, Нагорный | 265 703 |
| 42 | Коньково, Обручевский | 265 633 |
| 43 | Преображенское | 265 089 |
| 44 | Бутырский, Тимирязевский | 264 201 |
| 45 | Головинский | 261 962 |
| 46 | Войковский, Коптево | 258 285 |
| 47 | Строгино | 257 072 |
| 48 | Куркино, Молжаниновский | 252 300 |
| 49 | Лефортово | 251 269 |
| 50 | Чертаново Северное, Чертаново Центральное | 250 166 |
| 51 | Левобережный, Ховрино | 244 738 |
| 52 | Северное Тушино, Южное Тушино | 244 324 |
| 53 | Богородское, Метрогородок | 242 640 |
| 54 | Бабушкинский, Южное Медведково | 240 836 |
| 55 | Марфино | 238 698 |
| 56 | Северное Медведково | 236 949 |
| 57 | Рязанский | 235 777 |
| 58 | Кузьминки, Текстильщики | 233 754 |
| 59 | Ясенево | 233 696 |
| 60 | Печатники | 233 565 |
| 61 | Солнцево | 233 365 |
| 62 | Северный… | 232 455 |
| 63 | Чертаново Южное | 231 041 |
| 64 | Восточное Измайлово, Измайлово | 230 266 |
| 65 | Перово | 229 566 |
| 66 | Отрадное | 229 058 |
| 67 | Митино | 228 793 |
| 68 | Орехово-Борисово С/Ю | 228 227 |
| 69 | Москворечье-Сабурово, Царицыно | 228 079 |
| 70 | Внуково, Ново-Переделкино | 224 155 |
| 71 | Северное Бутово | 223 285 |
| 72 | Гольяново, Северное Измайлово | 222 655 |
| 73 | Лианозово | 222 472 |
| 74 | Капотня, Марьино | 222 314 |
| 75 | Алтуфьевский, Бибирево | 222 248 |
| 76 | Теплый Стан | 222 128 |
| 77 | Люблино | 218 060 |
| 78 | Ивановское, Новогиреево | 217 128 |
| 79 | Братеево, Зябликово | 216 985 |
| 80 | Лосиноостровский, Ярославский | 215 822 |
| 81 | Вешняки, Выхино-Жулебино | 210 302 |
| 82 | Восточный, Новокосино | 208 266 |
| 83 | Бирюлево Восточное, Бирюлево Западное | 203 134 |
| 84 | Южное Бутово | 202 134 |
| 85 | Жулебино | 197 933 |
| 86 | Зеленоград | 190 606 |

Средняя цена за 1 кв. м жилой площади характеризует стоимость жилой недвижимости в определенном районе г. Москвы.

Тип числовых данных – числовые (дискретные) данные (целые числа).

Количество наблюдений – 86.

Для анализа наличия в данных выбросов построим диаграмму типа «Ящик с усами» на Рисунке 4.

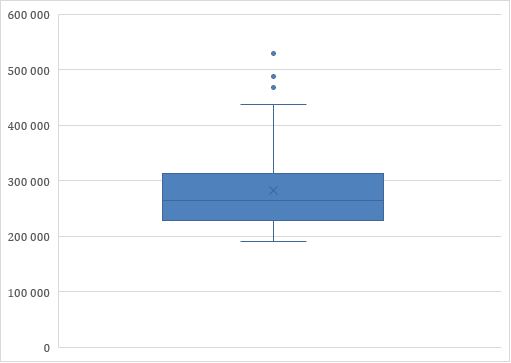


Рисунок 4. Диаграмма типа «Ящик с усами»

В верхней части графика видно наличие 3-х выбросов – точки со значениями 529 798 (Остоженка), 488 403 (Якиманка), 468 699 (Центр Москвы). При исключении данных наблюдений из анализа – появляются дополнительных 3 выброса – 491 032 (Арбат), 471 041 (Тверской), 436 960 (Хамовники). Причина выбросов – особая престижность районов благодаря близости к центру города и ключевым историческим достопримечательностям.

Итого необходимо исключить 6 наблюдений с максимальной стоимостью 1 кв.м. (Таблица 7).

Таблица 7. Скорректированные данные по средней цене 1 кв.м. в районах Москвы

| № | Район | Цена 1 кв. м. |
| --- | --- | --- |
| 1 | Пресненский | 391 878 |
| 2 | Мещанский | 391 147 |
| 3 | Замоскворечье | 390 349 |
| 4 | Красносельский | 381 260 |
| 5 | Беговой | 375 676 |
| 6 | Таганский | 374 585 |
| 7 | Дорогомилово | 366 909 |
| 8 | Гагаринский, Ломоносовский, Раменки | 353 391 |
| 9 | Донской | 351 960 |
| 10 | Хорошевский | 325 261 |
| 11 | Аэропорт, Сокол | 320 253 |
| 12 | Алексеевский | 315 348 |
| 13 | Марьина роща, Савеловский | 314 779 |
| 14 | Филевский парк | 314 468 |
| 15 | Проспект Вернадского | 314 197 |
| 16 | Басманный | 313 096 |
| 17 | Академический | 307 157 |
| 18 | Динамо | 306 647 |
| 19 | Сокольники | 306 595 |
| 20 | Черемушки | 304 532 |
| 21 | Хорошево-Мневники | 304 042 |
| 22 | Южнопортовый | 302 435 |
| 23 | Нижегородский | 296 601 |
| 24 | Крылатское | 289 663 |
| 25 | Тропарево-Никулино | 278 883 |
| 26 | Свиблово | 278 678 |
| 27 | Можайский, Фили-Давыдково | 278 512 |
| 28 | Даниловский, Котловка | 277 068 |
| 29 | Кунцево | 277 056 |
| 30 | Останкинский, Ростокино | 276 486 |
| 31 | Покровское-Стрешнево, Щукино | 273 189 |
| 32 | Очаково-Матвеевское | 272 741 |
| 33 | Соколиная гора | 272 303 |
| 34 | Нагатино-Садовники, Нагатинский Затон | 271 690 |
| 35 | Зюзино, Нагорный | 265 703 |
| 36 | Коньково, Обручевский | 265 633 |
| 37 | Преображенское | 265 089 |
| 38 | Бутырский, Тимирязевский | 264 201 |
| 39 | Головинский | 261 962 |
| 40 | Войковский, Коптево | 258 285 |
| 41 | Строгино | 257 072 |
| 42 | Куркино, Молжаниновский | 252 300 |
| 43 | Лефортово | 251 269 |
| 44 | Чертаново Северное, Чертаново Центральное | 250 166 |
| 45 | Левобережный, Ховрино | 244 738 |
| 46 | Северное Тушино, Южное Тушино | 244 324 |
| 47 | Богородское, Метрогородок | 242 640 |
| 48 | Бабушкинский, Южное Медведково | 240 836 |
| 49 | Марфино | 238 698 |
| 50 | Северное Медведково | 236 949 |
| 51 | Рязанский | 235 777 |
| 52 | Кузьминки, Текстильщики | 233 754 |
| 53 | Ясенево | 233 696 |
| 54 | Печатники | 233 565 |
| 55 | Солнцево | 233 365 |
| 56 | Северный… | 232 455 |
| 57 | Чертаново Южное | 231 041 |
| 58 | Восточное Измайлово, Измайлово | 230 266 |
| 59 | Перово | 229 566 |
| 60 | Отрадное | 229 058 |
| 61 | Митино | 228 793 |
| 62 | Орехово-Борисово С/Ю | 228 227 |
| 63 | Москворечье-Сабурово, Царицыно | 228 079 |
| 64 | Внуково, Ново-Переделкино | 224 155 |
| 65 | Северное Бутово | 223 285 |
| 66 | Гольяново, Северное Измайлово | 222 655 |
| 67 | Лианозово | 222 472 |
| 68 | Капотня, Марьино | 222 314 |
| 69 | Алтуфьевский, Бибирево | 222 248 |
| 70 | Теплый Стан | 222 128 |
| 71 | Люблино | 218 060 |
| 72 | Ивановское, Новогиреево | 217 128 |
| 73 | Братеево, Зябликово | 216 985 |
| 74 | Лосиноостровский, Ярославский | 215 822 |
| 75 | Вешняки, Выхино-Жулебино | 210 302 |
| 76 | Восточный, Новокосино | 208 266 |
| 77 | Бирюлево Восточное, Бирюлево Западное | 203 134 |
| 78 | Южное Бутово | 202 134 |
| 79 | Жулебино | 197 933 |
| 80 | Зеленоград | 190 606 |

В результате можно сделать выводы, что в данных Таблицы 2 выбросы отсутствуют (Рисунок 5).

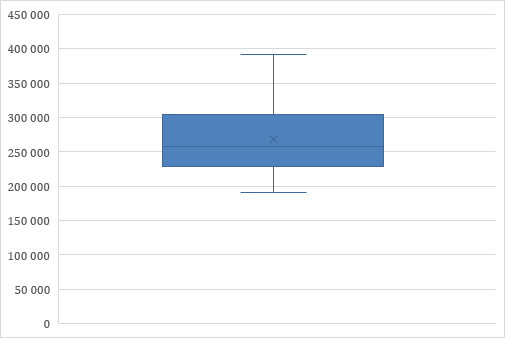


Рисунок 5. Диаграмма типа «Ящик с усами» после удаления выбросов

# ЛАБОРАТОРНАЯ РАБОТА №2

*1. Найти на сайте РосСтат непрерывное распределение случайной величины, которое было бы близко к нормальному. Построить графики распределения.*

Выбранные данные для анализа – уровень рождаемости за 2000-2022 гг. в России – представлены в Таблице 8 и на Рисунке 6.

Таблица 8. Уровень рождаемости в России за 2000-2022 гг[[5]](#footnote-4).

| Год | Уровень рождаемости |
| --- | --- |
| 2000 | 9,8 |
| 2001 | 10,0 |
| 2002 | 10,5 |
| 2003 | 11,1 |
| 2004 | 11,2 |
| 2005 | 11,0 |
| 2006 | 11,4 |
| 2007 | 12,9 |
| 2008 | 13,7 |
| 2009 | 13,9 |
| 2010 | 14,0 |
| 2011 | 14,1 |
| 2012 | 14,7 |
| 2013 | 14,5 |
| 2014 | 14,4 |
| 2015 | 12,8 |
| 2016 | 12,2 |
| 2017 | 11,2 |
| 2018 | 10,7 |
| 2019 | 9,8 |
| 2020 | 9,6 |
| 2021 | 9,5 |
| 2022 | 8,8 |

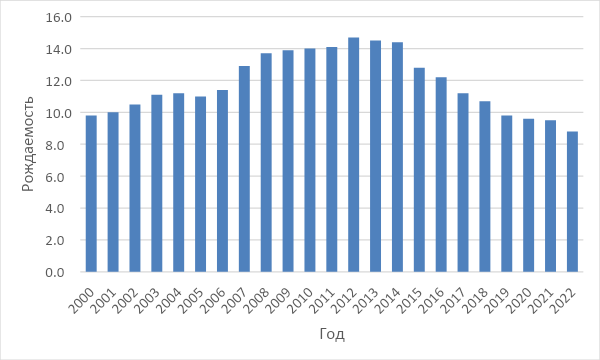


Рисунок 6. Уровень рождаемости в России за 2000-2022 гг.

*2. Посчитать в распределение близком к нормальному 1 л.р. и найденном непрерывном распределении:*

* *мат. ожидание*
* *2 (дисперсию), 3 (асимметрию), 4 (эксцесс) моменты*
* *квантили уровня 0,05 и 0,95; 2,5%-ную точку для найденной случайной величины.*

Для распределения из л.р.№1

Симметричным является распределение, в котором частоты любых двух вариантов, равностоящих в обе стороны от центра распределения, равны между собой.

Наиболее точным и распространенным показателем асимметрии является моментный коэффициент асимметрии.

где  - центральный момент третьего порядка; - среднеквадратическое отклонение.

= 32068327484219

=1,506

Положительная величина указывает на наличие правосторонней асимметрии

Оценка существенности показателя асимметрии дается с помощью средней квадратической ошибки коэффициента асимметрии:

=

Расчет центральных моментов представлен в Таблице 9.

Таблица 9. Расчет центральных моментов для распределения №1

| Группы | Середина интервала, xцентр | Кол-во, fi |  |  |
| --- | --- | --- | --- | --- |
| 0 - 7000 | 3500 | 4 | -1,270017638497E+14 | 4,2045900227251E+18 |
| 7000.1 - 10000 | 8500.05 | 6 | -1,2433925173364E+14 | 3,494740961258E+18 |
| 10000.1 - 14000 | 12000.05 | 10 | -1,4302787097173E+14 | 3,5194150567672E+18 |
| 14000.1 - 19000 | 16500.05 | 13 | -1,0404453914272E+14 | 2,0919713441952E+18 |
| 19000.1 - 27000 | 23000.05 | 18 | -45091236150703 | 6,1353382577488E+17 |
| 27000.1 - 45000 | 36000.05 | 21 | -4729601141,326 | 2868494815412 |
| 45000.1 - 60000 | 52500.05 | 15 | 59016862429406 | 9,3798460630127E+17 |
| 60000.1 - 75000 | 67500.05 | 6 | 1,7101311096528E+14 | 5,2831938428788E+18 |
| 75000.1 - 100000 | 87500.05 | 5 | 6,1956212279652E+14 | 3,1531685980779E+19 |
| 100000.1 - 150000 | 125000.05 | 4 | 2,9007500436803E+15 | 2,5640745406237E+20 |
| Итого |  | 100 | 3,2068327484219E+15 | 3,0808457257154E+20 |

==0,579

В анализируемом ряду распределения наблюдается несущественная асимметрия (1,506/0,579 = 2,6<3)

Чаще всего эксцесс оценивается с помощью показателя:

Ex=-3= 5,2223-3=2,22

Чтобы оценить существенность эксцесса рассчитывают статистику

*Ex* /Ex

где Ex - средняя квадратическая ошибка коэффициента эксцесса.

Ex===0,755

*Ex* /Ex= 2,22/0,755 = 2,941.

Поскольку значение < 3, то отклонение от нормального распределения считается не существенным.

Для найденного непрерывного распределения

Симметричным является распределение, в котором частоты любых двух вариантов, равностоящих в обе стороны от центра распределения, равны между собой.

Наиболее точным и распространенным показателем асимметрии является моментный коэффициент асимметрии.

где  - центральный момент третьего порядка; - среднеквадратическое отклонение.

=0,97

=0,155

Положительная величина указывает на наличие правосторонней асимметрии

Оценка существенности показателя асимметрии дается с помощью средней квадратической ошибки коэффициента асимметрии:

=

Расчет центральных моментов представлен в Таблице X

Таблица 10. Расчет центральных моментов

| x |  |  |  |
| --- | --- | --- | --- |
| 8,8 | 9,105 | -27,472 | 82,895 |
| 9,5 | 5,37 | -12,445 | 28,84 |
| 9,6 | 4,917 | -10,903 | 24,175 |
| 9,8 | 4,07 | -8,211 | 16,564 |
| 9,8 | 4,07 | -8,211 | 16,564 |
| 10 | 3,303 | -6,003 | 10,909 |
| 10,5 | 1,736 | -2,286 | 3,012 |
| 10,7 | 1,249 | -1,395 | 1,559 |
| 11 | 0,668 | -0,546 | 0,446 |
| 11,1 | 0,515 | -0,369 | 0,265 |
| 11,2 | 0,381 | -0,235 | 0,145 |
| 11,2 | 0,381 | -0,235 | 0,145 |
| 11,4 | 0,174 | -0,0727 | 0,0304 |
| 12,2 | 0,146 | 0,056 | 0,0214 |
| 12,8 | 0,966 | 0,949 | 0,932 |
| 12,9 | 1,172 | 1,269 | 1,374 |
| 13,7 | 3,544 | 6,672 | 12,561 |
| 13,9 | 4,337 | 9,033 | 18,812 |
| 14 | 4,764 | 10,397 | 22,694 |
| 14,1 | 5,21 | 11,893 | 27,147 |
| 14,4 | 6,67 | 17,226 | 44,487 |
| 14,5 | 7,196 | 19,305 | 51,788 |
| 14,7 | 8,309 | 23,953 | 69,047 |
| Итого | 78,253 | 22,369 | 434,413 |

==0,449

В анализируемом ряду распределения наблюдается несущественная асимметрия (0,155/0,449 = 0,34<3)

Чаще всего эксцесс оценивается с помощью показателя:

Ex=-3

=18,89

*Ex=*-3=-1,37

*Ex* < 0 - плосковершинное распределение

Чтобы оценить существенность эксцесса рассчитывают статистику

*Ex* /Ex

где Ex - средняя квадратическая ошибка коэффициента эксцесса.

Ex===0,744

*Ex* /Ex= -1,37/0,661 = 0,94

Поскольку значение < 3, то отклонение от нормального распределения считается не существенным.

*3. Построить сравнительную диаграмму найденного распределения и распределения по нормальному закону с параметрами.*

Сравнительная диаграмма представлена на Рисунке 7.

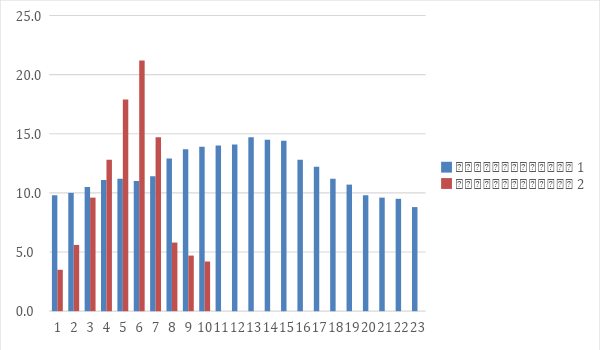


Рисунок 7. Сравнительная диаграмма

# ЛАБОРАТОРНАЯ РАБОТА №3

## Часть 1

*Найти на сайте РосСтат предположительно зависимые данные, посчитать коэффициент корреляции, оценить меру связи.*

*Построить гистограммы рассеяния, линию аппроксимации и посчитать величину достоверности аппроксимации R^2.*

Выбранные для анализа данные – размер ВВП в текущих ценах (X) и Среднемесячный уровень заработной платы (Y) за 2011-2022 гг. -представлены в Таблице 11.

Таблица 11. Динамика значений ВВП и среднемесячного уровня заработной платы за 2011-2022 гг.

| Год | ВВП (млрд. руб.)[[6]](#footnote-5) | Среднемесячный уровень з/п, руб[[7]](#footnote-6). |
| --- | --- | --- |
| 2011 | 60 114,0 | 23 369 |
| 2012 | 68 103,4 | 26 629 |
| 2013 | 72 985,7 | 29 792 |
| 2014 | 79 030,0 | 32 495 |
| 2015 | 83 087,4 | 34 030 |
| 2016 | 85 616,1 | 36 709 |
| 2017 | 91 843,2 | 39 167 |
| 2018 | 103 861,7 | 43 724 |
| 2019 | 109 608,3 | 47 867 |
| 2020 | 107 658,1 | 51 344 |
| 2021 | 135 773,8 | 57 244 |
| 2022 | 155 350,4 | 65 338 |

Диаграмма рассеяния, линия аппроксимации (линейного тренда) и уровень ее достоверности (R2) представлены на Рисунке 8.

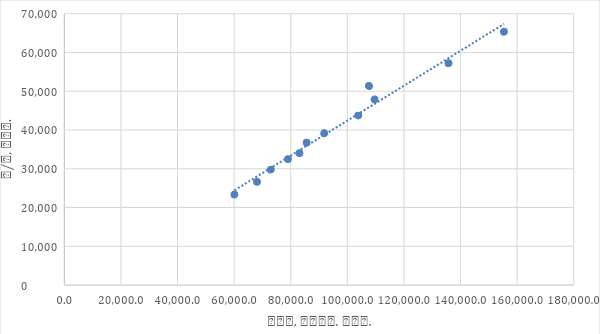


Рисунок 8. Диаграмма рассеяния и линейный тренд

Таким образом, зависимая переменная Y на 97,66% может быть объяснена в линейной модели y = 0,4507x - 2665,2 переменной Х, что является очень высоким уровнем надежности модели.

## Часть 2

*Найти датасет с объемом ~100 наблюдений (и более), сделать выборку =~ 1/3 наблюдений 1) случайным образом, 2) сделать стратифицированную выборку (разбив выборку на группы по смысловому содержанию, например, для регионов стратами могут быть федеральные округа).*

Исходный датасет– численность рабочей силы по регионам за 2019-2020 гг. – представлены в Таблице 12[[8]](#footnote-7).

Таблица 12. Численность рабочей силы по регионам за 2019-2020 гг.

| Субъект РФ | Численность рабочей силы, тыс. человек | |
| --- | --- | --- |
| 2019 | 2020 |
| Белгородская область | 826 | 834 |
| Брянская область | 595 | 583 |
| Владимирская область | 721 | 710 |
| Воронежская область | 1182 | 1172 |
| Ивановская область | 517 | 514 |
| Калужская область | 537 | 539 |
| Костромская область | 310 | 310 |
| Курская область | 569 | 557 |
| Липецкая область | 598 | 596 |
| Московская область | 4189 | 4154 |
| Орловская область | 347 | 350 |
| Рязанская область | 535 | 517 |
| Смоленская область | 483 | 477 |
| Тамбовская область | 500 | 499 |
| Тверская область | 676 | 656 |
| Тульская область | 793 | 792 |
| Ярославская область | 650 | 650 |
| г. Москва | 7308 | 7322 |
| Республика Карелия | 305 | 299 |
| Республика Коми | 428 | 415 |
| Архангельская область | 558 | 545 |
| в том числе: |  |  |
| Ненецкий автономный округ | 23 | 22 |
| Архангельская область без автономного округа | 535 | 523 |
| Вологодская область | 566 | 570 |
| Калининградская область | 537 | 531 |
| Ленинградская область | 969 | 981 |
| Мурманская область | 421 | 411 |
| Новгородская область | 303 | 288 |
| Псковская область | 316 | 302 |
| г. Санкт-Петербург | 3073 | 3097 |
| Республика Адыгея | 201 | 202 |
| Республика Калмыкия | 133 | 135 |
| Республика Крым | 918 | 925 |
| Краснодарский край | 2807 | 2821 |
| Астраханская область | 507 | 503 |
| Волгоградская область | 1243 | 1247 |
| Ростовская область | 2097 | 2111 |
| г. Севастополь | 222 | 217 |
| Республика Дагестан | 1 382 | 1287 |
| Республика Ингушетия | 259 | 261 |
| Кабардино-Балкарская Республика | 446 | 453 |
| Карачаево-Черкесская Республика | 204 | 204 |
| Республика Северная Осетия – Алания | 311 | 285 |
| Чеченская Республика | 634 | 653 |
| Ставропольский край | 1 383 | 1371 |
| Республика Башкортостан | 1 896 | 1901 |
| Республика Марий Эл | 334 | 326 |
| Республика Мордовия | 438 | 405 |
| Республика Татарстан | 2 036 | 2026 |
| Удмуртская Республика | 763 | 770 |
| Чувашская Республика | 608 | 604 |
| Пермский край | 1 225 | 1232 |
| Кировская область | 637 | 632 |
| Нижегородская область | 1 754 | 1737 |
| Оренбургская область | 930 | 933 |
| Пензенская область | 656 | 640 |
| Самарская область | 1 683 | 1676 |
| Саратовская область | 1 203 | 1164 |
| Ульяновская область | 611 | 599 |
| Курганская область | 365 | 371 |
| Свердловская область | 2 125 | 2109 |
| Тюменская область | 1 957 | 1952 |
| в том числе: |  |  |
| Ханты-Мансийский автономный округ – Югра | 915 | 912 |
| Ямало-Ненецкий автономный округ | 315 | 309 |
| Тюменская область без автономных округов | 727 | 731 |
| Челябинская область | 1 875 | 1855 |
| Республика Алтай | 95 | 99 |
| Республика Тыва | 117 | 131 |
| Республика Хакасия | 246 | 241 |
| Алтайский край | 1 140 | 1099 |
| Красноярский край | 1 482 | 1461 |
| Иркутская область | 1 167 | 1153 |
| Кемеровская область | 1 291 | 1276 |
| Новосибирская область | 1 430 | 1394 |
| Омская область | 1 016 | 1021 |
| Томская область | 541 | 537 |
| Республика Бурятия | 433 | 427 |
| Республика Саха (Якутия) | 501 | 502 |
| Забайкальский край | 525 | 524 |
| Камчатский край | 182 | 182 |
| Приморский край | 999 | 1002 |
| Хабаровский край | 702 | 721 |
| Амурская область | 408 | 403 |
| Магаданская область | 86 | 86 |
| Сахалинская область | 275 | 276 |
| Еврейская автономная область | 77 | 78 |
| Чукотский автономный округ | 31 | 31 |

Случайная выборка (50 наблюдений) – данные по регионам за 2020 г. – представлена в Таблице 13.

Таблица 13. Случайная выборка

| Субъект РФ | Численность рабочей силы,тыс. чел. |
| --- | --- |
| Алтайский край | 1099 |
| Амурская область | 403 |
| Архангельская область | 545 |
| Астраханская область | 503 |
| Белгородская область | 834 |
| Брянская область | 583 |
| Владимирская область | 710 |
| Волгоградская область | 1247 |
| Вологодская область | 570 |
| Воронежская область | 1172 |
| Еврейская автономная область | 78 |
| Забайкальский край | 524 |
| Ивановская область | 514 |
| Иркутская область | 1153 |
| Кабардино-Балкарская Республика | 453 |
| Калининградская область | 531 |
| Калужская область | 539 |
| Камчатский край | 182 |
| Карачаево-Черкесская Республика | 204 |
| Кемеровская область | 1276 |
| Кировская область | 632 |
| Костромская область | 310 |
| Краснодарский край | 2821 |
| Красноярский край | 1461 |
| Курганская область | 371 |
| Курская область | 557 |
| Ленинградская область | 981 |
| Липецкая область | 596 |
| Магаданская область | 86 |
| Московская область | 4154 |
| Мурманская область | 411 |
| Нижегородская область | 1737 |
| Новгородская область | 288 |
| Новосибирская область | 1394 |
| Омская область | 1021 |
| Оренбургская область | 933 |
| Орловская область | 350 |
| Пензенская область | 640 |
| Пермский край | 1232 |
| Приморский край | 1002 |
| Псковская область | 302 |
| Республика Адыгея | 202 |
| Республика Алтай | 99 |
| Республика Башкортостан | 1901 |
| Республика Бурятия | 427 |
| Республика Дагестан | 1287 |
| Республика Ингушетия | 261 |
| Республика Калмыкия | 135 |
| Республика Карелия | 299 |
| Республика Коми | 415 |

Стратифицированная выборка (те же регионы, но с разбивкой по федеральным округам) представлена в Таблице 14.

Таблица 14. Стратифицированная выборка

| Субъект РФ | Федеральный округ | Численность рабочей силы,тыс. чел. |
| --- | --- | --- |
| Амурская область | Дальневосточный федеральный округ | 403 |
| Еврейская автономная область | Дальневосточный федеральный округ | 78 |
| Забайкальский край | Дальневосточный федеральный округ | 524 |
| Камчатский край | Дальневосточный федеральный округ | 182 |
| Магаданская область | Дальневосточный федеральный округ | 86 |
| Приморский край | Дальневосточный федеральный округ | 1002 |
| Республика Бурятия | Дальневосточный федеральный округ | 427 |
| Кировская область | Приволжский федеральный округ | 632 |
| Нижегородская область | Приволжский федеральный округ | 1737 |
| Оренбургская область | Приволжский федеральный округ | 933 |
| Пензенская область | Приволжский федеральный округ | 640 |
| Пермский край | Приволжский федеральный округ | 1232 |
| Республика Башкортостан | Приволжский федеральный округ | 1901 |
| Архангельская область | Северо-Западный федеральный округ | 545 |
| Вологодская область | Северо-Западный федеральный округ | 570 |
| Калининградская область | Северо-Западный федеральный округ | 531 |
| Ленинградская область | Северо-Западный федеральный округ | 981 |
| Мурманская область | Северо-Западный федеральный округ | 411 |
| Новгородская область | Северо-Западный федеральный округ | 288 |
| Псковская область | Северо-Западный федеральный округ | 302 |
| Республика Карелия | Северо-Западный федеральный округ | 299 |
| Республика Коми | Северо-Западный федеральный округ | 415 |
| Кабардино-Балкарская Республика | Северо-Кавказский федеральный округ | 453 |
| Карачаево-Черкесская Республика | Северо-Кавказский федеральный округ | 204 |
| Республика Дагестан | Северо-Кавказский федеральный округ | 1287 |
| Республика Ингушетия | Северо-Кавказский федеральный округ | 261 |
| Алтайский край | Сибирский федеральный округ | 1099 |
| Иркутская область | Сибирский федеральный округ | 1153 |
| Кемеровская область | Сибирский федеральный округ | 1276 |
| Красноярский край | Сибирский федеральный округ | 1461 |
| Новосибирская область | Сибирский федеральный округ | 1394 |
| Омская область | Сибирский федеральный округ | 1021 |
| Республика Алтай | Сибирский федеральный округ | 99 |
| Курганская область | Уральский федеральный округ | 371 |
| Белгородская область | Центральный федеральный округ | 834 |
| Брянская область | Центральный федеральный округ | 583 |
| Владимирская область | Центральный федеральный округ | 710 |
| Воронежская область | Центральный федеральный округ | 1172 |
| Ивановская область | Центральный федеральный округ | 514 |
| Калужская область | Центральный федеральный округ | 539 |
| Костромская область | Центральный федеральный округ | 310 |
| Курская область | Центральный федеральный округ | 557 |
| Липецкая область | Центральный федеральный округ | 596 |
| Московская область | Центральный федеральный округ | 4154 |
| Орловская область | Центральный федеральный округ | 350 |
| Астраханская область | Южный федеральный округ | 503 |
| Волгоградская область | Южный федеральный округ | 1247 |
| Краснодарский край | Южный федеральный округ | 2821 |
| Республика Адыгея | Южный федеральный округ | 202 |
| Республика Калмыкия | Южный федеральный округ | 135 |

*Посчитать*

*1) среднее значение по выборкам случайной и стратифицированной,*

Среднее значение случайно выборки составляет =441,9 тыс. чел.

Для расчета среднего значения для стратифицированной выборки составим вспомогательную Таблицу 15.

Таблица 15. Вспомогательная таблица для стратифицированной выборки

| Федеральный округ | Число субъектов | Численность рабочей силы,тыс. чел. |
| --- | --- | --- |
| Дальневосточный федеральный округ | 7 | 1700 |
| Приволжский федеральный округ | 6 | 2205 |
| Северо-Западный федеральный округ | 9 | 4342 |
| Северо-Кавказский федеральный округ | 4 | 918 |
| Сибирский федеральный округ | 7 | 99 |
| Уральский федеральный округ | 1 | 371 |
| Центральный федеральный округ | 11 | 4993 |
| Южный федеральный округ | 5 | 840 |
| ИТОГО | 50 | 15 468 |

Таким образом, =309,36 (тыс. чел.)

*2) доверительный интервал для среднего (на уровне доверия 90%, 95%, 99%).*

Доверительный интервал для генерального среднего:

(

n=50; = 225,8 (расчет в MS Excel).

Уровень доверия 90%

Поскольку n>30, то определяем значение по таблицам функции Лапласа.

В этом случае 2Ф() = γ

Ф() = γ/2 = 0,90/2 = 0,45

По таблице функции Лапласа найдем, при каком значение Ф() = 0.45

(γ) = (0.45) = 1.65

(441,9 - 52,684;441,9 + 52,684) = (389,22;494,58)

Уровень доверия 95%

Ф() = γ/2 = 0.95/2 = 0.475

По таблице функции Лапласа найдем, при каком значение Ф() = 0,475

(γ) = (0,475) = 1,96

(441,9 - 62,582;441,9 + 62,582) = (379,32;504,48)

Уровень доверия 99%

Ф() = γ/2 = 0,99/2 = 0,495

По таблице функции Лапласа найдем, при каком tkp значение Ф() = 0,495

(γ) = (0.495) = 2.58

(441,9 - 82,378;441,9 + 82,378) = (359,52;524,28)

*Сравнить среднее генеральной выборки п.2 с полученными в выборках 2.1) и 2.2) и с границами доверительных интервалов из п.3.2).*

Среднее генеральной выборки составит 469,99 (тыс. чел.), среднее в случайной выборке: 441,9 (тыс. чел.), в стратифицированной: 309,36 (тыс. чел.). Границы доверительных интервалов: при уровне доверия 90% (389,22;494,58), при уровне доверия 95% (379,32;504,48), при уровне доверия 99% (359,52;524,28).

# СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Браки по возрастам жениха и невесты//Росстат. URL: https://rosstat.gov.ru/storage/mediabank/demo33\_2022.xls (режим доступа от 14.01.2024)
2. ВВП по годам//Росстат. URL: https://rosstat.gov.ru/storage/mediabank/VVP\_god\_S\_1995.xlsx (режим доступа от 14.01.2024)
3. Распределение общего объема денежных доходов по 20-ти процентным группам населения по Российской Федерации//Росстат. URL: https://rosstat.gov.ru/storage/mediabank/urov-32.xlsx (режим доступа от 14.01.2024)
4. Распределение общей суммы начисленной заработной платы по 10-процентным группам работников организаций (без субъектов малого предпринимательства) https://rosstat.gov.ru/storage/mediabank/raspr2\_2023.xls (режим доступа от 14.01.2024)
5. Рейтинг районов и метро Москвы по ценам на квартиры//Индикаторы рынка недвижимости. URL: https://www.irn.ru/kvartiry/moskva/ceny-po-rayonam/ (доступ от 14.01.2024 г.)
6. Рождаемость, смертность и естественный прирост//Росстат. URL: https://rosstat.gov.ru/storage/mediabank/demo21\_2022.xls (доступ от 14.01.2024)
7. Средняя начисленная заработная плата мужчин и женщин по обследованным видам экономической деятельности//Росстат. URL: https://rosstat.gov.ru/storage/mediabank/sr-zpl5.xlsx (режим доступа от 14.01.2024)
8. Численность и состав рабочей силы в возрасте 15 лет и старше//Росстат. URL: https://rosstat.gov.ru/storage/mediabank/Trud-1\_15-s.xlsx (режим доступа от 14.01.2024).

1. Распределение общего объема денежных доходов по 20-ти процентным группам населения по Российской Федерации//Росстат. URL: <https://rosstat.gov.ru/storage/mediabank/urov-32.xlsx> (режим доступа от 14.01.2024) [↑](#footnote-ref-0)
2. Распределение общей суммы начисленной заработной платы по 10-процентным группам работников организаций (без субъектов малого предпринимательства) <https://rosstat.gov.ru/storage/mediabank/raspr2_2023.xls> (режим доступа от 14.01.2024) [↑](#footnote-ref-1)
3. Браки по возрастам жениха и невесты//Росстат. URL: <https://rosstat.gov.ru/storage/mediabank/demo33_2022.xls> (режим доступа от 14.01.2024) [↑](#footnote-ref-2)
4. Рейтинг районов и метро Москвы по ценам на квартиры//Индикаторы рынка недвижимости. URL: <https://www.irn.ru/kvartiry/moskva/ceny-po-rayonam/> (доступ от 14.01.2024 г.) [↑](#footnote-ref-3)
5. Рождаемость, смертность и естественный прирост//Росстат. URL: <https://rosstat.gov.ru/storage/mediabank/demo21_2022.xls> (доступ от 14.01.2024) [↑](#footnote-ref-4)
6. ВВП по годам//Росстат. URL: <https://rosstat.gov.ru/storage/mediabank/VVP_god_S_1995.xlsx> (режим доступа от 14.01.2024) [↑](#footnote-ref-5)
7. Средняя начисленная заработная плата мужчин и женщин по обследованным видам экономической деятельности//Росстат. URL: <https://rosstat.gov.ru/storage/mediabank/sr-zpl5.xlsx> (режим доступа от 14.01.2024) [↑](#footnote-ref-6)
8. Численность и состав рабочей силы в возрасте 15 лет и старше//Росстат. URL: <https://rosstat.gov.ru/storage/mediabank/Trud-1_15-s.xlsx> (режим доступа от 14.01.2024). [↑](#footnote-ref-7)