

Введение

Курс «Статистика» имеет целью дать студентам представление о содержании статистики как научной дисциплины, познакомить с ее основными понятиями, методологией и методиками расчета важнейших статистических показателей.

Слово «статистика» имеет латинское происхождение (от слова status – состояние). Развитие статистической науки, расширение сферы практической статистической работы привели к изменению содержания самого понятия «статистика». В настоящее время это понятие употребляется в трех значениях:

- под статистикой понимают отрасль практической деятельности, которая имеет своей целью сбор, обработку, анализ и публикацию массовых данных о самых различных явлениях общественной жизни (в этом смысле «статистика» выступает как синоним словосочетания «статистический учет»);
- статистикой называют цифровой материал, служащий для характеристики какой-либо области общественных явлений;
- статистикой называется отрасль знания, особая научная дисциплина и соответственно учебный предмет в вузах.

Как и всякая наука «статистика» имеет свой предмет, которым является количественная характеристика массовых общественных явлений в неразрывной связи и их количественной характеристикой. Свой предмет «статистика» изучает при помощи определенных категорий, т.е. понятий, которые отражают наиболее общие и существенные свойства, признаки, связи и отношения между ними.

В дисциплине «Статистика» излагаются вопросы общей теории статистики: предмет, метод статистики как общественной науки, ее значение; способы получения статистической информации, основы ее научной разработки, правила и приемы сравнения данных; виды средних и их применение в статистико-экономическом анализе; показатели вариации; методы изучения временных рядов; индексный метод анализа; основы регрессионного и корреляционного анализа.

В соответствии с учебным планом, рабочей программой курса «Статистика» студенты выполняют домашнее задание по теме: «Статистическое изучение связи между признаками», вариант которого определяется последней цифрой студенческого билета.

I. Исходные данные для выполнения домашнего задания

В результате выборочного обследования 10% рабочих авиаремонтного завода (по состоянию на 1 января текущего года) получены следующие данные:

№№ п\п	Разряд	Производственный стаж, лет	Заработная плата, у.е.
1	2	3	4
Цех № 1			
1	4	5	539
2	1	1	487
3	4	7	554
4	2	2	507
5	1	1	490
6	2	5	519
7	3	8	536
8	5	10	574
9	2	0	481
10	3	7	533
11	2	2	515
12	2	3	524
13	5	5	553
14	1	1	479
15	3	4	509
16	3	8	552
17	2	3	526
18	2	1	495
19	1	0	492
20	4	6	562
21	2	5	516
22	1	0	483
23	4	8	531
24	4	12	548
25	2	4	521
26	3	7	529
27	3	6	520
28	2	1	475
29	3	8	525
30	1	0	472
31	4	3	553
32	2	4	518
33	1	0	485
34	2	3	508

№№ п\п	Разряд	Производственный стаж, лет	Зарботная плата, у.е.
35	3	8	507
36	5	17	578
37	2	1	505
38	6	23	600
39	3	4	528
40	3	11	538
Цех № 2			
1	3	5	536
2	2	1	501
3	3	3	517
4	4	15	571
5	2	1	492
6	4	19	562
7	1	0	480
8	3	5	541
9	3	7	535
10	2	1	502
11	3	3	528
12	4	12	565
13	4	2	525
14	5	6	536
15	5	8	574
16	3	3	523
17	6	29	571
18	2	3	498
19	4	13	537
20	3	8	530
21	1	1	494
22	2	0	468
23	4	3	513
24	3	9	547
25	6	9	594
26	5	12	588
27	1	2	504
28	3	6	523
29	1	0	460
30	4	14	536
31	2	4	517
32	3	5	535
33	3	0	492
34	4	15	553

№№ п\п	Разряд	Производственный стаж, лет	Заработная плата, у.е.
35	5	8	573
36	2	1	486
37	4	2	543
38	3	4	522
39	3	7	534
40	4	10	558
41	2	4	506
42	2	4	512
43	3	11	552
44	4	5	527
45	4	7	547
46	5	15	595
47	3	4	514
48	3	8	555
49	3	9	524
50	2	4	505
51	4	11	559
52	1	1	491
53	3	9	534
54	4	10	552
55	3	2	526
56	5	21	597
57	3	8	521
58	2	0	483
59	5	13	575
60	2	2	508

Вариант № 1

1. Построить ряд распределения рабочих каждого цеха и всего завода по разрядам.

2. Определить количественную взаимосвязь между признаками:

2.1. С помощью графического метода определить форму связи между разрядом и заработной платой рабочих цеха № 1 с № 1 по № 20 включительно ($n=20$).

2.2. Вычислить параметры уравнения регрессии, характеризующего зависимость между разрядом рабочих и их заработной платой. Построить на графике теоретическую и эмпирическую линии регрессии. Объяснить смысл полученных параметров уравнения.

2.3. Определить степень тесноты между рассматриваемыми признаками.

Вариант № 2

1. Построить ряд распределения рабочих каждого цеха и всего завода по размеру заработной платы, выделив семь групп с равными интервалами.
2. Определить количественную взаимосвязь между признаками:
 - 2.1. С помощью графического метода определить форму связи между тарифным разрядом и заработной платой рабочих цеха № 1 с № 21 по № 40 включительно ($n=20$).
 - 2.2. Вычислить параметры уравнения регрессии, характеризующего зависимость между разрядом рабочих и их заработной платой. Построить на графике теоретическую и эмпирическую линии регрессии. Объяснить смысл полученных параметров уравнения.
 - 2.3. Определить степень тесноты между рассматриваемыми признаками

Вариант № 3

1. Построить ряд распределения рабочих завода по общему стажу работы, выделив пять групп со следующими специализированными интервалами: 1) менее года; 2) от 1 до 2 лет; 3) от 3 до 5 лет; 4) от 6 до 10 лет; 5) от 11 лет и выше.
2. Определить количественную взаимосвязь между признаками:
 - 2.1. С помощью графического метода определить форму связи между производственным стажем и заработной платой рабочих цеха № 1 с № 1 по № 20 включительно ($n=20$).
 - 2.2. Вычислить параметры уравнения регрессии, характеризующего зависимость между производственным стажем рабочих и их заработной платой. Построить на графике теоретическую и эмпирическую линии регрессии. Объяснить смысл полученных параметров уравнения.
 - 2.3. Определить степень тесноты между рассматриваемыми признаками

Вариант № 4

1. Произвести комбинационное распределение рабочих по разрядам и размеру заработной платы.
2. Определить количественную взаимосвязь между признаками:
 - 2.1. С помощью графического метода определить форму связи между тарифным разрядом и заработной платой рабочих цеха № 2 с № 1 по № 20 включительно ($n=20$).
 - 2.2. Вычислить параметры уравнения регрессии, характеризующего зависимость между тарифным разрядом и заработной платой рабочих.

Построить на графике теоретическую и эмпирическую линии регрессии. Объяснить смысл полученных параметров уравнения.

2.3. Определить степень тесноты между рассматриваемыми признаками

Вариант № 5

1. Построить комбинационное распределение рабочих завода по общему стажу работы и тарифным разрядам.

2. Определить количественную взаимосвязь между признаками:

2.1. С помощью графического метода определить форму связи между тарифным разрядом и заработной платой рабочих цеха № 2 с № 21 по № 40 включительно ($n=20$).

2.2. Вычислить параметры уравнения регрессии, характеризующего зависимость между тарифным разрядом и заработной платой рабочих. Построить на графике теоретическую и эмпирическую линии регрессии. Объяснить смысл полученных параметров уравнения.

2.3. Определить степень тесноты между рассматриваемыми признаками

Вариант № 6

1. Построить комбинационное распределение рабочих завода по общему стажу работы и заработной плате.

2. Определить количественную взаимосвязь между признаками:

2.1. С помощью графического метода определить форму связи между тарифным разрядом и заработной платой рабочих цеха № 2 с № 41 по № 60 включительно ($n=20$).

2.2. Вычислить параметры уравнения регрессии, характеризующего зависимость между тарифным разрядом и заработной платой рабочих. Построить на графике теоретическую и эмпирическую линии регрессии. Объяснить смысл полученных параметров уравнения.

2.3. Определить степень тесноты между рассматриваемыми признаками

Вариант № 7

1. Построить ряд распределения рабочих завода по размеру заработной платы, выделив 7 групп с равными интервалами. Определить в целом по заводу моду заработной платы рабочих.

2. Определить количественную взаимосвязь между признаками:

2.1. С помощью графического метода определить форму связи между производственным стажем и заработной платой рабочих цеха № 1 с № 21 по № 40 включительно ($n=20$).

2.2. Вычислить параметры уравнения регрессии, характеризующего зависимость между производственным стажем и заработной платой рабочих. Построить на графике теоретическую и эмпирическую линии регрессии. Объяснить смысл полученных параметров уравнения.

2.4. Определить степень тесноты между рассматриваемыми признаками

Вариант № 8

1. Построить ряд распределения рабочих завода по разрядам. Определить в целом по заводу моду разряда рабочих.

2. Определить количественную взаимосвязь между признаками:

2.1. С помощью графического метода определить форму связи между производственным стажем и заработной платой рабочих цеха № 2 с № 1 по № 20 включительно ($n=20$).

2.2. Вычислить параметры уравнения регрессии, характеризующего зависимость между производственным стажем и заработной платой рабочих. Построить на графике теоретическую и эмпирическую линии регрессии. Объяснить смысл полученных параметров уравнения.

2.3. Определить степень тесноты между рассматриваемыми признаками.

Вариант № 9

1. Построить ряд распределения рабочих завода по заработной плате и определить ее медиану.

2. Определить количественную взаимосвязь между признаками:

2.1. С помощью графического метода определить форму связи между производственным стажем и заработной платой рабочих цеха № 2 с № 21 по № 40 включительно ($n=20$).

2.2. Вычислить параметры уравнения регрессии, характеризующего зависимость между производственным стажем и заработной платой рабочих. Построить на графике теоретическую и эмпирическую линии регрессии. Объяснить смысл полученных параметров уравнения.

2.3. Определить степень тесноты между рассматриваемыми признаками.

Вариант № 0

1. Построить ряд распределения рабочих завода по стажу работы, выделив пять групп со следующими специализированными интервалами:

1) менее года; 2) от 1 до 2 лет; 3) от 3 до 5 лет; 4) от 6 до 10 лет; 5) от 11 и выше. Определить моду производственного стажа рабочих.

2. Определить количественную взаимосвязь между признаками:

2.1. С помощью графического метода определить форму связи между производственным стажем и заработной платой рабочих цеха № 2 с № 41 по № 60 включительно ($n=20$).

2.2. Вычислить параметры уравнения регрессии, характеризующего зависимость между производственным стажем и заработной платой рабочих. Построить на графике теоретическую и эмпирическую линии регрессии. Объяснить смысл полученных параметров уравнения.

2.3. Определить степень тесноты между рассматриваемыми признаками.

II. Методические указания по выполнению домашнего задания

Домашнее задание «Статистическое изучение связи между признаками» имеет своей целью закрепить и проверить знания студентов, полученные ими в процессе изучения теоретического материала по следующим темам:

- сводка и группировка статистических данных;
- статистическое изучение связи между признаками.

Тема «Сводка и группировка данных статистического наблюдения»

Собранный в процессе статистического наблюдения материал представляет собой разрозненные первичные сведения об отдельных единицах изучаемого явления. В таком виде материал еще не характеризует явление в целом: не дает представления ни о величине (численности) явления, ни о его составе, ни о размере характерных признаков, ни о существовании связей этого явления с другими явлениями и т.д. Указанные сведения нельзя получить непосредственно из наблюдения. Они могут быть выявлены лишь в процессе обработки материалов наблюдения. Началом такой обработки и служит сводка и группировка данных наблюдения.

Основным и важнейшим моментом сводки является группировка, т.е. расчленение статистической совокупности на группы и подгруппы по определенным существенным признакам. Выбор группировочного признака - один из самых сложных вопросов теории группировок. После выбора группировочного признака определяется число групп, если оно не указано в условии задания. Определяя число групп, нужно иметь в виду, что в каждую группу должно попасть также число единиц совокупности, на основании которого можно делать обоснованные выводы.

В результате группировки единиц совокупности по какому-либо варьирующему признаку получают ряды распределения.

Статистическим рядом распределения называют упорядоченное распределение единиц совокупности на группы по изучаемому признаку. В нем различают следующие элементы: варианты (x) и частоты (f) или частости (f'). Вариантами называют отдельные значения группировочного признака, которые он принимает в ряду распределения. Числа, которые показывают сколько раз (как часто) встречается в ряду распределения то или иное значение варианты, называют частотами. Частость - это относительный показатель структуры совокупности, т.е. он характеризует долю частот отдельных вариантов в общей сумме частот. Сумма всех частостей равна единице. Частости могут выражаться и в процентах, тогда сумма всех частостей равна 100%.

Ряды распределения могут быть построены как по количественному признаку, так и по атрибутивному (не имеющего количественного разряда: профессия, пол, отрасль и т.д.). В соответствии с этим ряды делятся на вариационные и атрибутивные. Вариационные ряды могут быть дискретными и интервальными.

Дискретный ряд распределения - это ряд, в котором варианты выражены одним конечным числом (например, группировка рабочих по разряду).

Интервальный ряд - это ряд, в котором значения признака выражены в виде интервала (например, группировка рабочих по уровню заработной платы, стажа).

Интервалом называется разность между максимальным и минимальным значением признака в каждой группе. Причем, интервалы бывают равные (если совокупность более или менее однородная), неравные (если разброс значений признака в совокупности большой), открытые (если нет начального или конечного значения признака в первой или последней группах) и закрытые.

Если строится интервальный ряд с равными интервалами и количество групп оговаривается исследованием, то величина интервала определяется следующим образом:

$$i = \frac{x_{MAX} - x_{MIN}}{K},$$

где i - величина интервала;

x_{MAX} - максимальное значение признака в совокупности;

x_{MIN} - минимальное значение признака;

K - количество принятых групп.

Если принимаются неравные интервалы, то следует, чтобы частоты по группам распределялись бы более менее равномерно.

Результаты группировки оформляются в виде статистической таблицы. Надо уметь правильно построить таблицу и проанализировать ее содержание.

При группировке по одному признаку строятся так называемые групповые таблицы (простой ряд распределения), а при группировке по нескольким признакам - комбинационные таблицы (комбинационное распределение).

Например, группировку рабочих по разряду можно представить в виде следующей таблицы:

Группировка рабочих по разряду (x)	Число рабочих (f)
1	3
2	7
3	10
4	6
5	8
6	6
Итого:	40

Группировку рабочих по заработной плате и разряду можно показать с помощью комбинационного распределения:

Группировка рабочих по заработной плате, у.е.	500 - 520	520 - 540	540 - 560	560 - 580	580 - 600	600 - 620	620 - 640	Итого
Группировка рабочих по разряду								
1	3							3
2		7						7
3			8	2				10
4				4	2			6
5					5	3		8
6						2	4	6
Итого:	3	7	8	6	7	5	4	40

Причем, при построении комбинационного распределения следует количество групп по результативному признаку следует брать больше, чем по факториальному признаку.

Тема «Статистическое изучение связи между признаками»

По характеру зависимости между факториальными и результативными признаками связи делятся на функциональные и стохастические.

Стохастическая связь называется корреляционной, если при изменении значений факториальных признаков меняется средняя величина результативного признака. Уравнение, характеризующее изменение средней величины результативного признака в зависимости от изменений значений факториального признака, называется уравнением корреляционной связи или уравнением регрессии. Корреляционный анализ предназначен для изучения тесноты связи между факториальным и результативным признаками, а регрессионный анализ - для нахождения уравнения корреляционной связи (регрессии), оценки ее точности и надежности.

Первая задача сводится к определению формы и вида связи. Существует несколько методов выявления наличия связи между признаками: метод параллельных рядов или параллельного сопоставления, графический метод (построение поля корреляции), способ группировки и выведение средних по группам.

Если уравнение связывает два признака (один факториальный и один результативный), то это - уравнение парной регрессии. При определении функции, связывающий результативный признак с одним факторным, используется графическое изображение связи. Полученная эмпирическая линия (ломаная) регрессии показывает, какую функцию для отображения связи можно применить.

В случае линейной парной регрессии уравнение имеет следующий вид:

$$\bar{y} = a + b x ,$$

где \bar{y} - среднее значение результативного признака;

x - значение факториального признака;

a, b - параметры уравнения связи.

Параметр (коэффициент регрессии) показывает, на сколько в среднем изменяется результативный признак при изменении факториального признака на единицу. Параметр характеризует изменение результативного признака от всех прочих факторов, которые здесь не учитываются.

Для определения параметров уравнения прямой на основе методов наименьших квадратов решается следующая система нормальных уравнений:

$$\begin{cases} \sum y = a \cdot n + b \sum x \\ \sum xy = a \cdot \sum x + b \sum x^2, \end{cases}$$

где n - объем совокупности;

y - индивидуальные значения результативного признака;

x - индивидуальные значения факториального признака.

При наличии линейной зависимости степень тесноты связи можно рассчитать с помощью коэффициента парной корреляции (τ)

Коэффициент парной корреляции можно определить следующим образом:

$$\tau = \frac{\sum xy - \frac{\sum x \cdot \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) \cdot \left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}.$$

Коэффициент корреляции изменяется в пределах от -1 до +1. Знак при коэффициенте корреляции говорит о направлении связи (прямая или обратная), а его величина - о степени тесноты связи (чем ближе к 1, тем связь теснее).

Список рекомендуемой литературы:

1. Теория статистики / под ред. проф. Л.Г. Громько. - М.: Инфра-М., 2011.
2. Громько Г.Л Теория статистики: практикум. - М.: Инфра-М., 2011.