СОДЕРЖАНИЕ

Введение

Общая теория статистики представляет собой науку, которая разрабатывает общие принципы и методы статистического исследования, используемые во всех отраслях, в том числе на воздушном транспорте.

Эти методы используются для:

- получения итоговых обобщающих показателей;
- изучения особенностей распределения единиц совокупности по тому или иному признаку;
- определения средней величины того или иного количественного показателя и его вариации;
- организации выборочных обследований и оценки точности полученных выборочных характеристик;
- выявления взаимосвязи между отдельными показателями;
- изучения динамики отдельных показателей;
- расчета экономических индексов.

В соответствии с учебным планом подготовки по направлению 190700 «Технология транспортных процессов» по дисциплине «Общая теория статистики» выполняется курсовая работа, целью которой является закрепление и проверка знаний студентов, полученных в процессе изучения теоретического материала, а также их умение использовать в практической деятельности.

Курсовая работа состоит из пяти разделов:

- в первом разделе требуется построить различные ряды распределения, где в основе группировки лежат один признак (простая группировка) или два признака, взятых в комбинации друг с другом (комбинационная группировка или комбинационное распределение);
- во втором разделе необходимо рассчитать относительные величины (относительные показатели структуры) или средние величины (среднюю арифметическую и структурные средние -моду и медину);
- третий раздел содержит расчет абсолютных показателей вариации (размах вариации, среднее линейное и квадратическое отклонение, дисперсия) и относительного показателя вариации (коэффициента вариации);
- в четвертом разделе следует определить показатели выборочного наблюдения (среднюю и предельную ошибки выборки), объем выборочной совокупности, вероятность, с которой можно гарантировать точность среднего значения того или иного показателя в генеральной совокупности на основе средней выборочной;
- в пятом разделе необходимо показать взаимосвязь между изучаемыми признаками (результативным и факториальным) в количественном выражении. Для этого следует определить форму и вид связи между признаками, рассчитать параметры уравнения связи, а также коэффициенты степени тесноты и существенности связи.

Вариант курсовой работы определяется последней цифрой зачетной книжки.

Курсовая работа выполняется на листах формата А4 с подробными пояснениями расчетов.

1. Исходные данные для выполнения курсовой работы

В результате выборочного обследования 10% рабочих авиаремонтного завода (по состоянию на 01 января текущего года) получены следующие данные:

№№ п\п	Разряд	Производственный стаж, полных 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
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
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

2. Варианты заданий на курсовую работу

Вариант № 1

1. Построить ряд распределения рабочих каждого цеха и всего завода по квалификации (разрядам). Рассчитать относительные величины структуры, характеризующие состав рабочих по квалификации.

2. Определить моду и медиану квалификации рабочих каждого цеха и завода в целом, а также медиану заработной платы рабочих завода.

3. Определить дисперсию тарифного разряда в каждом цехе и по заводу в целом; среднюю из цеховых дисперсий; межцеховую дисперсию. Объяснить смысл дисперсий. Используя их, проверить правило сложения дисперсий.

4. С вероятностью 0,954 определить ошибку выборки для среднего тарифного разряда рабочих завода и для доли рабочих, имеющих пятый разряд. Указать пределы возможных значений этих показателей в генеральной совокупности. Какая должна быть численность выборки, чтобы ошибка выборки с этой вероятностью для среднего тарифного разряда не превысила 0,2?

5. Определить количественную связь между признаками:

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

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

5.3. Определить степень тесноты связи 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 моду и медиану заработной платы рабочих всего завода.

3. Определить дисперсию заработной платы рабочих в каждом цехе и по заводу в целом; среднюю из цеховых дисперсий; межцеховую дисперсию. Объяснить смысл дисперсий. Используя их, проверить правило сложения дисперсий.

4. С вероятностью 0,954 определить ошибку выборки для средней заработной платы рабочих завода и для доли рабочих, имеющих заработную плату менее 500 у.е. Указать пределы возможных значений этих показателей в генеральной совокупности. Какая вероятность того, что доля рабочих, имеющих заработную плату до 500 у.е., в генеральной совокупности не превышает 25%?

5. Определить количественную взаимосвязь между признаками:

5.1. С помощью графического метода определить форму связи между тарифным разрядом и заработной платой рабочих цеха № 1 с № 21 по № 40 включительно (п=20).

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

5.3. 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 моду и медиану стажа рабочих всего завода.

3. Определить дисперсию производственного стажа рабочих в каждом цехе и по заводу в целом; среднюю из цеховых дисперсий; межцеховую дисперсию. Объяснить смысл дисперсий. Используя их, проверить правило сложения дисперсий.

4. С вероятностью 0,954 определить ошибку выборки для среднего производственного стажа рабочих завода и для доли рабочих, имеющих стаж менее 5 лет. Указать пределы возможных значений этих показателей в генеральной совокупности. Какая должна быть численность выборки, чтобы ошибка этой выборки с этой вероятностью для производственного стажа не превысила 0,5 года?

5. Определить количественную взаимосвязь между признаками:

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

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

5.3.Определить степень тесноты между рассматриваемыми признаками и существенность связи.

Вариант № 4

1. Произвести комбинационное распределение рабочих каждого цеха и завода в целом по разряду и размеру месячной заработной платы.

2. Вычислить по данным комбинационного распределения рабочих всего завода среднюю заработную плату каждой группы по квалификации и среднюю заработную плату по всем группам вместе взятым.

3. Рассчитать дисперсию заработной платы в цехе № 1 и № 2. Определить коэффициенты вариации заработной платы рабочих по цехам. Сделать выводы.

4. С вероятностью 0,997 определить ошибку выборки для средней заработной платы рабочих завода и для доли рабочих, имеющих заработную плату более 500 у.е. Указать пределы возможных значений этих показателей в генеральной совокупности. Какая вероятность того, что доля рабочих, имеющих заработную плату свыше 500 у.е., в генеральной совокупности не превысит 85%?

5. Определить количественную взаимосвязь между признаками:

5.1. С помощью графического метода определить форму связи между тарифным разрядом и заработной платой рабочих цеха № 2 с № 1 по № 20 включительно (n=20).

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

5.3.Определить степень тесноты связи между рассматриваемыми признаками и существенность связи.

Вариант № 5

1. Построить комбинационное распределение рабочих каждого цеха и завода в целом по общему стажу работы и разрядам.

2. По данным комбинационного распределения рабочих завода исчислить средний разряд каждой группы рабочих по общему стажу и средний разряд рабочих по всем группам вместе взятым.

3. Рассчитать дисперсию производственного стажа рабочих в цехе № 1 и № 2. Определить коэффициенты вариации производственного стажа рабочих по цехам. Сделать выводы.

4. С вероятностью 0,954 определить ошибку выборки для доли рабочих, имеющих производственный стаж более 10 лет. Согласуются ли выборочные данные с предположением о том, что доля рабочих, имеющих производственный стаж более 10 лет, в генеральной совокупности составляет 5%, 10%, 15%, 20%? Выводы сделать с той же вероятностью.

5. Определить количественную взаимосвязь между признаками:

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

5.2. 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.

5.3. Определить степень тесноты между рассматриваемыми признаками.

Вариант № 6

1. Построить комбинационное распределение рабочих каждого цеха и завода в целом по общему стажу работы и заработной плате.

2. Рассчитать средние тарифный разряд, заработную плату и производственный стаж рабочих цеха № 2, а также моду и медиану заработной платы этих рабочих.

3. Рассчитать дисперсию тарифного разряда рабочих в цехах № 1 и № 2. Определить коэффициенты вариации тарифного разряда рабочих по цехам. Сделать выводы.

4. С вероятностью 0,954 определить ошибку выборки для среднего тарифного разряда рабочих завода и для доли рабочих, имеющих четвертый разряд. Указать пределы возможных значений этих показателей в генеральной совокупности.

5. Определить количественную взаимосвязь между признаками:

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

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

5.3. Определить степень тесноты между рассматриваемыми признаками

Вариант № 7

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

2. Рассчитать средний тарифный разряд, средний производственный стаж рабочих завода, коэффициент вариации этих показателей. Сделать выводы.

3. Рассчитать среднюю заработную плату и дисперсию заработной платы рабочих завода обычным способом и способом условных моментов.

4. С вероятностью 0,954 определить ошибку выборки средней заработной платы рабочих цеха № 1 и для доли рабочих цеха № 1, имеющих заработную плату менее 500 у.е. Указать пределы возможных значений этих показателей в генеральной совокупности. Какая вероятность того, что доля рабочих, имеющих заработную плату до 500 у.е., в генеральной совокупности не превысит 30%?

5. Определить количественную взаимосвязь между признаками:

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

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

5.3.Определить степень тесноты связи между рассматриваемыми признаками

Вариант № 8

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

2. Определить среднюю заработную плату рабочих цеха № 1 обычным способом и способом условных моментов.

3. Рассчитать абсолютные и относительные показатели вариации тарифного разряда рабочих по цехам. Сделать выводы.

4. С вероятностью 0,954 определить ошибку выборки для среднего тарифного разряда рабочих цеха № 1 и для доли рабочих цеха № 1, имеющих третий разряд. Указать пределы возможных значений этих показателей в генеральной совокупности. Какая должна быть численность выборки, чтобы ошибка выборки с этой вероятностью для среднего тарифного разряда не превысила 0,2?

5. Определить количественную взаимосвязь между признаками:

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

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

5.3. Определить степень тесноты между рассматриваемыми признаками.

Вариант № 9

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

2. На основе комбинационного распределения исчислить среднюю заработную плату каждой группы рабочих по стажу и среднюю заработную плату рабочих по всем группам вместе взятым.

3. Рассчитать абсолютные и относительные показатели вариации заработной платы рабочих по цехам. Сделать выводы.

4. С вероятностью 0,954 определить ошибку выборки для средней заработной платы рабочих цеха № 2 и для доли рабочих этого цеха, имеющих заработную плату менее 500 у.е. Указать пределы возможных значений этих показателей в генеральной совокупности. Какая вероятность того, что доля рабочих цеха № 2, имеющих заработную плату до 500 у.е., в генеральной совокупности не превысит 25%?

5. Определить количественную взаимосвязь между признаками:

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

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

5.3. Определить степень тесноты между рассматриваемыми признаками и существенность связи.

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 и медиану производственного стажа рабочих.

2. Определить среднюю заработную плату рабочих цеха № 2 обычным способом и способом условных моментов.

3. Рассчитать абсолютные и относительные показатели вариации заработной платы рабочих по цехам. Сделать выводы

4. С вероятностью 0,954 определить ошибку выборки средней заработной платы рабочих цеха № 1 и для доли рабочих цеха № 1, имеющих заработную плату менее 500 у.е. Указать пределы возможных значений этих показателей в генеральной совокупности.

5. Определить количественную взаимосвязь между признаками:

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

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

5.3. Определить степень тесноты между рассматриваемыми признаками.

3. Методические указания по выполнению курсовой работы

3.1. 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 представление ни о величине (численности) явления, ни о его составе, ни о размере характерных признаков, ни о существе 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 такое 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^I).

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 вариант 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 значения: 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 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 зарплате, у.е. / Группировка рабочих по разряду	500-520	520-540	540-560	560-580	580-600	Итого
1	3					3
2		7				7
3			8	2		10
4				4	2	6
5			1	2	5	8
6				2	4	6
Итого:	3	7	9	10	11	40

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

3.2. Средние величины

Средняя величина – это обобщенная характеристика единиц совокупности по определенному признаку. В статистике используются различные виды средних величин: арифметическая, гармоническая, структурные средние: мода и медиана и др.

Все виды средних могут быть рассчитаны как по индивидуальным значениям осредняемого принципа, так и по сгруппированным. В первом случае вычисления средняя будет называться простой средней, а во втором – средней арифметической взвешенной.

Средняя арифметическая простая определяется по формуле ($\bar{x}$):

$$\bar{x} = \frac{\sum x}{n},$$

где x - индивидуальные значения признака (варианты);

n - объем совокупности.

Средняя арифметическая взвешенная вычисляется по формуле:

$$\bar{x} = \frac{\sum x \cdot f}{\sum f},$$

где f - частота повторения признака.

Средняя арифметическая в отличие от других средних обладает рядом математических свойств. Их необходимо изучить, так как на основании этих свойств среднюю арифметическую можно вычислить упрощенным способом (способом условных моментов).

Для характеристики структуры вариационных рядов применяются показатели особого рода. К ним относят моду и медиану.

Мода (M_o) – это значение варьирующего признака наиболее часто встречающееся в данном ряду.

Модой в дискретном ряду является варианта, имеющая наибольшую частоту. В интервальном ряду с равными интервалами моду определяют следующим образом:

$$M_O = x_{Mo} + i\,\frac{f_{Mo} - f_{Mo-1}}{(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})},$$

где x_{Mo} - нижняя граница модального интервала;

f_{Mo} - частота модального интервала, т.е. интервала, содержащего моду;

f_{Mo-1} - частота интервала, предшествующая модальному;

f_{Mo+1} - частота интервала следующего за модальным.

Медиана (M_e) – это варианта, которая находится в середине упорядоченного вариационного ряда, т.е. она делит ряд на две равные части.

Если ряд распределения дискретный и имеет нечетное число членов, то медианой будет варианта, находящаяся в середине данного ряда. Если упорядоченный ряд состоит из четного числа членов, то медиана будет определяться как средняя арифметическая из двух вариант, расположенных в середине ряда. Если же в дискретном ряду значения признаков повторяются, то для расчета медианы используется сумма накопленных частот. Наращивание частот должно продолжаться до получения суммы частот немногим больше половины. Варианта, соответствующая этой величине, и будет медианой. Если сумма накопленных частот против одной из вариант будет равна точно половине суммы частот, то медиана определяется как средняя арифметическая из этой варианты и последующей.

В интервальном ряду распределения с равными интервалами медиана определяется по формуле:

$$M_E = x_{Me} + i\,\frac{\frac{1}{2}\sum f - S_{Me-1}}{f_{Me}}\ ,$$

где x_{Me} - нижняя граница медианного интервала;

f_{Me} - частота медианного интервала;

S_{Mo-1}- сумма накопленных частот, предшествующая медианному интервалу.

3.3. Показатели вариации

Для характеристики колеблемости (вариации) изучаемого признака в совокупности используются специальные показатели вариации к которым относятся:

- размах вариации (R);
- среднее линейное отклонение ($\overline{d}$);
- дисперсия (σ^2);
- среднее квадратичное отклонение (σ);
- коэффициент вариации (V).

Размах вариации представляет собой разность между наибольшим и наименьшим значениями варьирующего признака:

$$R = x_{MAX} - x_{MIN}\,.$$

Среднее линейное отклонение представляет собой среднюю арифметическую из абсолютных отклонений вариант от их среднего значения. Среднее линейное отклонение может быть простым и взвешенным:

$$\bar{d} = \frac{\sum |x - \bar{x}|}{n}; \quad \bar{d} = \frac{\sum |x - \bar{x}| \cdot f}{\sum f}.$$

Дисперсия представляет собой среднюю арифметическую из квадратов отклонений вариант от их средней арифметической. Дисперсия вычисляется по формулам простой и взвешенной средней:

$$\sigma^2 = \frac{\sum (x - \bar{x})^2}{n}; \quad \sigma^2 = \frac{\sum (x - \bar{x})^2 \cdot f}{\sum f}.$$

Среднее квадратическое отклонение представляет собой корень квадратный из дисперсии:

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}; \quad \sigma = \sqrt{\frac{\sum (x - \bar{x})^2 \cdot f}{\sum f}}.$$

Вышеперечисленные показатели относятся к абсолютным показателям вариации.

Относительным показателем вариации является коэффициент вариации, который определяется следующим образом:

$$V = \frac{\sigma}{\bar{x}} \cdot 100\%.$$

Коэффициент вариации является критерием надежности средней: если он велик (более 40 %), то это свидетельствует о большой колеблемости в величине признака у отдельных единиц данной группы, а, следовательно, средняя недостаточно надежна. Таким образом, коэффициент вариации является показателем степени однородности совокупности.

В том случае, если совокупность разбита на группы, то помимо общей дисперсии, которая рассчитывается для всей совокупности в целом, можно рассчитать групповые дисперсии (σ_i^2) и межгрупповую (σ^2).

Между этими дисперсиями существует связь:

$$\sigma^2 = \overline{\sigma_i^2} + \partial^2,$$

которая называется правилом сложения дисперсий, где $\overline{\sigma_i^2}$ - средняя из групповых дисперсий.

$$\sigma_i^2 = \frac{\sum (x - \bar{x}_i)^2 \cdot f}{\sum f}; \quad \overline{\sigma_i^2} = \frac{\sum \sigma_i^2 f}{\sum f}$$

$\bar{x}_i$ - среднее значение признака в группе.

Межгрупповая дисперсия рассчитывается следующим образом

$$\partial^2 = \frac{\sum(\overline{x}_i - \overline{x}_o)^2 \cdot f}{\sum f}; \ \partial^2 = \frac{\sum(\overline{x}_i - \overline{x}_o)^2}{n}$$

где $\overline{x}_o$ - средняя для всей совокупности.

Правило сложения дисперсий используется для определения степени тесноты связи между признаками (результативным и факториальным) с помощью эмпирического корреляционного отношения (η), которое определяется по следующей формуле:

$$\eta = \sqrt{\frac{\partial^2}{\sigma^2}}.$$

Причем, эмпирическое корреляционное отношение рассчитывается по результативному признаку.

В статистике разработаны приемы упрощенного вычисления средней арифметической и дисперсии. Одним из наиболее эффективных способов упрощения является способ условных моментов, который основан на свойствах средней арифметической. Используя его, получают следующие формулы:

$$\overline{x} = A + i \cdot M_1;$$

$$\sigma^2 = i^2(M_2 - M_1^2);$$

$$\sigma = i\sqrt{M_2 - M_1^2},$$

где А – условное начало (обычно за условное начало принимают варианту, наиболее часто встречающуюся в совокупности);

M_1 – условный момент первого порядка, который определяется:

$$M_1 = \frac{\sum\left(\frac{x - A}{i}\right) \cdot f}{\sum f};$$

M_2 – условный момент второго порядка

$$M_2 = \frac{\sum\left(\frac{x - A}{i}\right)^2 \cdot f}{\sum f}$$

3.4. Показатели выборочного наблюдения

Выборочным называется такое наблюдение, при котором характеристика всей совокупности единиц дается по некоторой ее части, отобранной в случайном порядке. Выборочное наблюдение – наиболее распространенный вид несплошного наблюдения. Оно дает возможность, не прибегая к сплошному наблюдению, получить обобщающие показатели, которые правильно отражают характеристики всей совокупности в целом.

Вся совокупность единиц называется генеральной совокупностью, а та часть совокупности единиц, которая подвергается выборочному обследованию, называется выборочной совокупностью. Задача выборочного наблюдения – получить правильное представление о показателях генеральной совокупности на основе изучения выборочной совокупности.

Основными вопросами теории выборочного наблюдения являются:

- определение предельной ошибки выборки для различных типов выборочных характеристик с учетом особенностей отбора;
- определение объема выборки, обеспечивающего необходимую репрезентативность выборочной совокупности с учетом особенностей отбора.

Величина предельной ошибки выборки зависит от вариации признака внутри совокупности объема выборки. И способа отбора единиц.

Средняя ошибка выборки, сформированной по количественному признаку определяется:

$$\mu_x = \sqrt{\frac{\sigma^2}{n}},$$

где σ^2 - дисперсия количественного признака;

n - объем выборочной совокупности.

Средняя ошибка выборки, сформированной по альтернативному признаку равно:

$$\mu_P = \sqrt{\frac{\tilde{p}\,(1-\tilde{p})}{n}} = \sqrt{\frac{\sigma_p^2}{n}};$$

где $\tilde{p}$ (1-$\tilde{p}$) – дисперсия альтернативного признака;

$\tilde{p}$ – среднее значение альтернативного признака, которое равно частоте его появления в выборке, т.е.

$$\tilde{p} = \frac{n_1}{n}$$

n_1 – часть выборочной совокупности, обладающая тем или иным признаком.

В математической статистике доказано, что с определенной степенью вероятности можно утверждать, что выборочные и генеральные характеристики не превысят заданной величины Δ, которая называется предельной ошибкой.

Предельная и средняя ошибки связаны между собой следующим образом:

$$\Delta = t \cdot \mu ,$$

где μ – средняя ошибка выборки;

t – коэффициент доверия, зависящий от вероятности (ρ), с которой можно утверждать, что предельная ошибка не превысит t – кратное значение средней ошибки.

Значения вероятности ρ от t устанавливаются математической статистикой. Их краткая выдержка из таблицы значений функции Лапласа при разных значениях t представлена ниже:

t	Ф(t)= ρ	t	Ф(t)= ρ
1.0	0.683	2.2	0.972
1.2	0.770	2.5	0.987
1.5	0.866	2.7	0.993
1.7	0.911	2.9	0.996
1.8	0.928	3.0	0.997
2.0	0.954	3.6	0.999

Определение ошибок выборочных характеристик позволяет установить границы нахождения соответствующих генеральных показателей.

Для средней: $\bar{x} = \tilde{x} \pm \Delta x$ - (в основе количественный признак).

Для доли: $\bar{p} = \tilde{p} \pm \Delta p$ - (в основе альтернативный признак),

где $\bar{x}$, $\bar{p}$ - показатели генеральной совокупности;

$\tilde{x}$, $\tilde{p}$ - показатели выборочной совокупности.

При расчете предельной ошибки выборки учитывается способ отбора единиц из генеральной совокупности.

Повторный способ отбора:

$$\Delta x = t \sqrt{\frac{\sigma_x^2}{n}}; \ \Delta p = t \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n}}$$

Бесповторный отбор:

$$\Delta x = t \sqrt{\frac{\sigma_x^2}{n} \cdot \left(1 - \frac{n}{N}\right)}; \ \Delta p = t \sqrt{\frac{\tilde{p}\left(1 - \tilde{p}\right)}{n} \cdot \left(1 - \frac{n}{N}\right)},$$

где N –объем генеральной совокупности;

$\frac{n}{N}$ - обследованная часть совокупности;

$\left(1-\frac{n}{N}\right)$- необследованная часть совокупности.

Как видно из формул ошибка выборки при бесповторном отборе будет меньше.

При организации выборочного наблюдения большое значение имеет правильное определение необходимой численности выборки. Формула объема выборки получается из соответствующей формулы предельной ошибки. Например, при определении средней:

$$n=\frac{t^2\cdot\sigma_X^2}{\Delta_X^2};\ n=\frac{t^2\cdot\sigma_X^2\cdot N}{\Delta_X^2 N+t^2\sigma_X^2}$$

при определении среднего значения в виде доли:

$$n=\frac{t^2\cdot\tilde{p}\left(1-\tilde{p}\right)}{\Delta_P^2};\ n=\frac{t^2\cdot\tilde{p}\left(1-\tilde{p}\right)\cdot N}{\Delta_P^2 N+t^2\tilde{p}\left(1-\tilde{p}\right)}$$

В практике выборочного наблюдения очень часто ставится вопрос о вероятности получения того или иного значения выборочной средней, не выходящего за известные пределы. В том случае может быть применена формула Гаусса, в которой:

$$t=\frac{\bar{x}-\tilde{x}}{\mu_X}\ \text{или}\ t=\frac{\bar{p}-\tilde{p}}{\mu_P},$$

т.е. в числителе этих отношений - заданное отклонение выборочной средней от средней генеральной совокупности, а в знаменателе – средняя ошибка.

3.5. Статистическое изучение взаимосвязей 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 факторным 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, связывающей результативный признак с одним факториальным, 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:

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

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

x - индивидуальные значения факториального признака;

$a\,,b$ - параметры уравнения связи.

Параметр «b» – (коэффициент регрессии) показывает на сколько в среднем изменяется результативный признак (y) при изменении факториального признака (x) на единицу. Параметр «а» - характеризует изменение результативного признака под воздействием всех прочих факторов, которые здесь не учитываются.

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

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

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) или эмпирического корреляционного отношения (η).

Коэффициент корреляции можно определить следующим образом:

$$\tau_{XY} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right)\left(\sum y^2 - \frac{\left(\sum y\right)^2}{n}\right)}};$$

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).

Оценка существенности коэффициента корреляции определяется на основании критерия его надежности t, который рассчитывается по формуле:

$$t = \frac{|\tau|\sqrt{n-1}}{1-\tau^2}$$

В математической статистике доказано, что если $t < 2{,}56$, то связь между признаками признается несущественной. В этом случае считается, что факториальный признак не оказывает существенного влияния на результативный признак. Если $t > 2{,}56$, то связь признается существенной, т.е. факториальный признак оказывает существенное влияние на признак результативный.

Литература

1. Теория статистики/ под ред. проф. Л.Г.Громыко.-М.:Инфра-М.,2011.
2. Громыко Г.Л.Теория статистики:практикум.-М.:Инфра-М.,2011.
3. Степанова Н.И. Статистика: учеб. пособие. - М.: МГТУГА, 2006. - Часть I.
4. Степанова Н.И. Статистика: пособие по проведению практических занятий. – М.: МГТУ ГА. 2008.

№	Программное обеспечение и интернет-ресурсы
5.	Электронные ресурсы библиотеки Университета – электронные версии пособий, методических разработок, указаний и рекомендаций по всем видам учебной работы.