

СТАТИСТИЧЕСКИЕ РАСПРЕДЕЛЕНИЯ И ИХ ОСНОВНЫЕ ХАРАКТЕРИСТИКИ

§ 21.1. ПОСТРОЕНИЕ РЯДОВ РАСПРЕДЕЛЕНИЯ

Изучение структуры той или иной совокупности достигается построением **рядов распределения**, характеризующих распределение единиц совокупности по одному признаку.

Ряды распределения делятся на вариационные и атрибутивные. **Вариационный ряд** — это распределение единиц совокупности по количественному признаку. В **атрибутивных рядах** представлена группировка по атрибутивным (качественным) признакам.

Форма построения вариационного ряда зависит от характера изменения изучаемого признака, он может быть построен в форме дискретного ряда или в форме интервального ряда.

По характеру вариации значений признака различают:

- ♦ признаки с прерывным изменением (дискретные);
- ♦ признаки с непрерывным изменением (непрерывные).

Признаки с **прерывным изменением** могут принимать лишь конечное число определенных значений (например, тарифный разряд работников, число туристов в группе и др.). Признаки с **непрерывным изменением** могут принимать в определенных границах любые значения (например, стаж работы, размер дохода и т. д.).

Для признака, имеющего прерывное изменение и принимающего небольшое количество значений, применяется построение **дискретного ряда**. В первой графе ряда указываются конкретные значения каждого индивидуального значения признака, во второй графе — численность единиц с определенным значением признака.

Для признака, имеющего непрерывное изменение, строится **интервальный вариационный ряд**, состоящий, так же как и дискретный ряд, из двух граф (варианты и частоты). При его построении в первой графе отдельные значения признака указываются в виде интервалов «от — до», во второй графе — число еди-

ниц, входящих в интервал. Интервалы **обращаются, как правило, равные и закрытые**.

Величина интервала определяется по формуле:

$$i = \frac{R}{m}, \quad R = x_{\max} - x_{\min},$$

где R — размах колебания (варьирования) признака; $x_{\max}$, $x_{\min}$ — максимальное и минимальное значения признака в совокупности соответственно; m — число групп.

Число групп приближенно определяется по формуле *Стерджесса*:

$$m = 1 + 3,322 \lg n,$$

где n — общее число единиц совокупности.

Поскольку количество групп не может быть дробным числом, то полученную по этой формуле величину округляют до целого большего числа.

При небольшом объеме информации число групп может быть установлено исследователем без использования формулы Стерджесса.

Величину интервала обычно округляют до целого (всегда большего) числа.

Нижнюю границу первого интервала принимают равной минимальному значению признака (чаще всего его предварительно округляют до целого меньшего числа); верхняя граница первого интервала соответствует значению $(x_{\min} + i)$. Для последующих групп границы определяются аналогично, то есть последовательно прибавляется величина интервала. Если единица обладает значением признака, равным величине верхней границы интервала, то ее следует относить к следующей группе.

Примером интервального вариационного ряда служит *табл. 21.1*.

Таблица 21.1. Группы туристских предприятий по объему выручки от реализации туристского продукта

Группы туристских предприятий по объему выручки от реализации туристского продукта, в тыс. руб.	Число предприятий	Накопленная частота, S
80—90	2	2
90—100	22	24
100—110	48	72
110—120	16	88
120—130	2	90
Итого	90	

В каждой выделенной группе различают нижнюю и верхнюю границы интервала. Так, в последней группе предприятий (*табл. 21.1*) нижняя граница — 120 тыс. руб., верхняя — 130 тыс. руб.

При построении атрибутивных рядов число групп соответствует числу разновидностей признака.

Ряд распределения, состоящий из двух граф (варианты и частоты), иногда дополняется другими графами, необходимыми для вычисления отдельных статистических показателей. Достаточно часто в ряд вводится графа, в которой подсчитываются **накопленные частоты** S (см. табл. 21.1). Накопленные частоты показывают, какое число единиц совокупности обладает значением признака «не более» или «не менее» определенного, и исчисляются путем последовательного прибавления к частоте первого интервала частот последующих интервалов.

Если вариационный ряд дан с неравными интервалами, то для правильного представления о характере распределения необходимо произвести расчет абсолютной или относительной плотности распределения.

Абсолютная плотность распределения (p) представляет собой величину частоты, приходящейся на единицу размера интервала отдельной группы ряда:

$$p = \frac{f}{i}.$$

Относительная плотность распределения (p') — это частное от деления частоты (w) отдельной группы на размер ее интервала:

$$p' = \frac{w}{i}.$$

Эти показатели используются для преобразования интервалов, что бывает необходимо при сравнительной оценке двух группировок.

Для наглядности вариационный ряд изображают графически. Дискретный вариационный ряд изображается с помощью **полигона** (или многоугольника) распределения частот, являющегося разновидностью статистических ломаных. Для изображения интервального ряда применяются **полигон распределения частот** и **гистограмма частот**.

Графики строятся в прямоугольной системе координат. При построении полигона частот на оси абсцисс в одинаковом масштабе откладываются направо в порядке возрастания значения признака (для дискретного характера) или центральные значения интервалов (для интервальных рядов); по оси ординат наносится шкала для выражения величин частот. Из точек на оси абсцисс, соответствующих величине признака, восстанавливаются перпендикуляры высотой, соответствующей частоте. Вершины перпендикуляров соединяются отрезками прямой. Крайние точки полученной ломаной соединяются с лежащими на оси абсцисс следующими

ми (меньшими и большими) возможными, но фактически не наблюдающимися значениями признака, частота которых, очевидно, равна 0. Замкнутая с осью абсцисс ломаная линия представляет полигон распределения частот.

Для построения гистограммы по оси абсцисс откладывают величины интервалов, а частоты изображаются прямоугольниками, построенными на интервалах с высотой в масштабе оси ординат. В случае неравенства интервалов гистограмма строится не по частотам или частостям, а по плотности распределения.

В ряде случаев для изображения вариационных рядов используется **кумулятивная кривая** (кумулята). Она особенно удобна для сравнения вариационных рядов. Накопленные частоты наносятся на чертеж в виде ординат. Соединяя вершины отдельных ординат прямыми, получают ломаную линию, которая, начиная с нуля, непрерывно поднимается над осью абсцисс до тех пор пока не достигнет высоты, соответствующей общей сумме частот.

Если поменять местами оси координат в кумуляте, то получаем новый вид графического изображения — **огиву**.

При изучении процессов концентрации (концентрации производства, концентрации капитала и др.) используется графическое изображение вариационного ряда в виде **кривой Лоренца** (или кривой концентрации). Для ее построения абсолютные показатели числа единиц в группах и размер изучаемого признака выражаются в относительных показателях (в долях или процентах к итогу) и исчисляются их накопленные значения.

При построении графика на горизонтальной линии наносится шкала для ряда накопленных частостей, а на вертикальной линии — шкала для накопленных относительных величин размера изучаемого признака. Далее наносятся точки в соответствии с накопленными значениями двух рядов. Соединив все точки прямыми линиями, получают кривую, характеризующую степень неравномерности распределения.

Линия, соединяющая нижний левый угол графика с верхним правым (диагональ четырехугольника), является линией равномерного распределения. Чем больше фактическое распределение двух показателей отклоняется от равномерного, тем больше кривая удалена от диагонали.

Если значения признака в вариационном ряду даны в порядке убывания (от большего к меньшему), то построенная по таким данным кривая Лоренца будет расположена выше диагонали в форме выпуклости.

Для анализа вариационных рядов используются три группы показателей:

- ♦ показатели центра распределения;
- ♦ показатели степени вариации;
- ♦ показатели формы распределения.

§ 21.2. ПОКАЗАТЕЛИ ЦЕНТРА РАСПРЕДЕЛЕНИЯ

Для характеристики среднего значения признака в вариационном ряду применяются: средняя арифметическая, медиана, мода.

Средняя арифметическая для дискретного ряда распределения вычисляется по формуле:

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

где x — варианты значений признака; f — частота повторения данного варианта.

Средняя арифметическая для интервального ряда распределения вычисляется по формуле:

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

где x' — середина соответствующего интервала значения признака (вычисляется как средняя из значений границ интервала).

Медиана (Me) соответствует варианту, стоящему в середине ранжированного ряда. Для нахождения медианы сначала определяется ее порядковый номер:

$$N_{Me} = \frac{n + 1}{2},$$

где n — число единиц в совокупности.

По накопленным частотам определяют ее численное значение в дискретном вариационном ряду.

Если совокупность содержит четное число значений варьирующего признака ($n = 2k$; $k = n/2$), то в этом случае за медиану условно принимают значение:

$$Me = \frac{1}{2}(x_k + x_{k+1}),$$

так как в ряду нет члена, который делил бы совокупность на две равные по объему группы.

В интервальном ряду распределения сначала указывают интервал, в котором находится медиана.

Медианным является первый интервал, в котором сумма накопленных частот превысит половину общего числа наблюдений.

Численное значение медианы определяется по формуле:

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

где x_{Me} — нижняя граница медианного интервала; i — величина интервала; S_{Me-1} — накопленная частота интервала, предшествующего медианному; f_{Me} — частота медианного интервала.

Мода (Mo) — наиболее часто встречающееся значение признака. В дискретном ряду — это варианта с наибольшей частотой. В интервальном ряду сначала определяется модальный интервал, то есть тот интервал, который имеет наибольшую частоту.

Конкретное значение моды определяется по формуле:

$$Mo = x_{Mo} + i \frac{f_{Mo} - f_{Mo-1}}{[f_{Mo} - f_{Mo-1}] + [f_{Mo} - f_{Mo+1}]},$$

где x_{Mo} — нижняя граница модального интервала; f_{Mo} — частота модального интервала; f_{Mo-1} — частота интервала, предшествующего модальному; f_{Mo+1} — частота интервала, следующего за модальным.

§ 21.3. ПОКАЗАТЕЛИ ВАРИАЦИИ (КОЛЕБЛЕМОСТИ) ПРИЗНАКА

Для суждения о размере вариации признака в статистике используются абсолютные и относительные показатели. К абсолютным показателям вариации относятся:

- ✧ размах колебаний;
- ✧ среднее линейное отклонение;
- ✧ среднее квадратическое отклонение;
- ✧ дисперсия;
- ✧ квартильное отклонение.

Размах колебаний (размах вариации) определяется по формуле:

$$R = x_{\max} - x_{\min},$$

где $x_{\max}$, $x_{\min}$ — максимальное и минимальное значения признака соответственно.

Величина показателя зависит от величины только двух крайних вариантов и не учитывает степени колеблемости основной массы членов ряда.

Среднее линейное отклонение ($\bar{d}$) и среднее квадратическое отклонение (σ) показывают, на сколько в среднем отличаются индивидуальные значения признака от среднего его значения.

Среднее линейное отклонение определяется по формулам:

а) для несгруппированных данных (первичного ряда):

$$\bar{d} = \frac{\sum |x - \bar{x}|}{n};$$

б) для вариационного ряда:

$$\bar{d} = \frac{\sum |x - \bar{x}|f}{\sum f}.$$

Среднее квадратическое отклонение (σ) и дисперсия (σ^2) определяются так:

а) для несгруппированных данных:

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

б) для вариационного ряда:

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

Формула для расчета дисперсии может быть преобразована:

$$\begin{aligned} \sigma^2 &= \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n [x_i^2 - 2x_i\bar{x} + (\bar{x})^2]}{n} = \\ &= \frac{\sum_{i=1}^n x_i^2 - 2\bar{x}\sum_{i=1}^n x_i + n(\bar{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - (\bar{x})^2. \end{aligned}$$

То есть дисперсия равна средней из квадратов индивидуальных значений признака минус квадрат средней величины. Следовательно,

$$\sigma^2 = \overline{x^2} - (\bar{x})^2.$$

Среднее квадратическое отклонение по своей величине всегда превышает значение среднего линейного отклонения в соответствии со свойством мажорантности средних.

Квартильное отклонение (d_k) применяется вместо размаха вариации, чтобы избежать недостатков, связанных с использованием крайних значений:

$$d_k = \frac{Q_3 - Q_1}{2},$$

где Q_3 и Q_1 — третья и первая квартили распределения соответственно.

Квартиль — это значения признака, которые делят ранжированный ряд на четыре равные по численности части. Таких величин будет три: первая квартиль (Q_1), вторая квартиль (Q_2), третья квартиль (Q_3). Вторая квартиль является медианой. Вычисление квартилей аналогично вычислению медианы.

Сначала определяют положение или место квартили:

$$N_{Q_1} = \frac{n+1}{4}; \quad N_{Q_2} = \frac{n+1}{4} \times 2 = \frac{n+1}{2}; \quad N_{Q_3} = \frac{n+1}{4} \times 3;$$

Затем по накопленным частотам в дискретном ряду определяют численное значение.

В интервальном ряду распределения сначала указывают интервал, в котором лежит квартиль, затем определяют ее численное значение по формуле:

$$Q = x_Q + i \frac{N_Q - S_{Q-1}}{f_Q},$$

где x_Q — нижняя граница интервала, в котором находится квартиль; S_{Q-1} — накопленная частота интервала, предшествующего тому, в котором находится квартиль; f_Q — частота интервала, в котором находится квартиль.

При сравнении колеблемости различных признаков в одной и той же совокупности или же при сравнении колеблемости одного и того же признака в нескольких совокупностях с различной величиной средней арифметической используются **относительные показатели вариации**. Они вычисляются как отношение абсолютных показателей вариации к средней арифметической (или медиане) и чаще всего выражаются в процентах. Используются следующие формулы расчета относительных показателей вариации:

♦ коэффициент осцилляции:

$$K_R = \frac{R}{\bar{x}} \times 100\%;$$

♦ относительное линейное отклонение:

$$K_{\bar{d}} = \frac{\bar{d}}{\bar{x}} \times 100\%;$$

♦ коэффициент вариации:

$$v = \frac{\sigma}{\bar{x}} \times 100\%;$$

♦ относительный показатель квартильной вариации:

$$K_{d_k} = \frac{d_k}{Me} \times 100\% \quad \text{или} \quad K_Q = \frac{Q_3 - Q_1}{2Q_2} \times 100\%.$$

Наиболее часто применяется коэффициент вариации. Его применяют не только для сравнительной оценки вариации, но и для характеристики однородности совокупности. Совокупность считается однородной, если коэффициент вариации не превышает 33% (для распределений, близких к нормальному).

§ 21.4. СЛОЖЕНИЕ ДИСПЕРСИЙ ИЗУЧАЕМОГО ПРИЗНАКА

Изучая дисперсию интересующего нас признака в пределах исследуемой совокупности и опираясь на общую среднюю в расчетах, нельзя оценить влияние отдельных факторов, определяющих

колеблемость индивидуальных значений (вариант) признака. Это можно сделать при помощи метода группировок, когда единицы изучаемой совокупности подразделяются на однородные группы по признаку-фактору. При этом кроме общей средней для всей совокупности исчисляются средние по отдельным группам (групповые или частные средние) и три показателя дисперсии:

- ♦ общая дисперсия;
- ♦ межгрупповая дисперсия;
- ♦ средняя внутригрупповая дисперсия.

Величина **общей дисперсии** (σ_0^2) характеризует вариацию признака под влиянием всех факторов, формирующих уровень признака у единиц данной совокупности, и определяется по формуле:

$$\sigma_0^2 = \frac{\sum (x - \bar{x}_0)^2 f}{\sum f},$$

где $\bar{x}_0$ — общая средняя арифметическая для всей изучаемой совокупности.

Межгрупповая дисперсия (дисперсия групповых средних δ^2) отражает систематическую вариацию, то есть те различия в величине изучаемого признака, которые возникают под влиянием фактора, положенного в основу группировки. Межгрупповая дисперсия определяется по формуле:

$$\delta^2 = \frac{\sum (\bar{x}_i - \bar{x}_0)^2 n_i}{\sum n_i},$$

где $\bar{x}_i$ — средняя по отдельной группе; n_i — число единиц в определенной группе.

Средняя внутригрупповая дисперсия ($\bar{\sigma}^2$) характеризует случайную вариацию, возникающую под влиянием других, неучтенных факторов, и не зависит от условия (признака-фактора), положенного в основу группировки. Средняя внутригрупповая дисперсия определяется по формуле:

$$\bar{\sigma}^2 = \frac{\sum \sigma_i^2 n_i}{\sum n_i},$$

где σ_i^2 — дисперсия по отдельной группе:

$$\sigma_i^2 = \frac{\sum (x - \bar{x}_i)^2 f}{\sum f}.$$

Указанные дисперсии взаимосвязаны между собой следующим равенством: величина общей дисперсии равна сумме межгрупповой дисперсии и средней внутригрупповой дисперсии:

$$\sigma_0^2 = \delta^2 + \bar{\sigma}^2.$$

Это тождество отражает закон (правило) сложения дисперсии. Опираясь на это правило, можно определить, какая часть (доля)

общей дисперсии складывается под влиянием признака-фактора, положенного в основу группировки.

§ 21.5. ВАРИАЦИИ АЛЬТЕРНАТИВНОГО ПРИЗНАКА

Альтернативный признак — качественный признак, имеющий две взаимоисключающие разновидности (например, работники предприятия подразделяются на мужчин и женщин; продукция — на годную и бракованную и т. д.).

Альтернативный признак принимает всего два значения: 1 — наличие признака; 0 — отсутствие признака:

$$p + q = 1,$$

где p — доли единиц, обладающих признаком; q — доли единиц, не обладающих признаком.

Среднее значение альтернативного признака определяется по формуле:

$$\bar{x} = \frac{(1 \times p) + (0 \times q)}{p + q} = p.$$

Дисперсия альтернативного признака вычисляется следующим образом:

$$\sigma^2 = \frac{(1 - p)^2 \times p + (0 - p)^2 \times q}{p + q} = p \times q.$$

Предельное значение вариации альтернативного признака равно 0,25; оно получается при $p = q = 0,5$.

§ 21.6. ПОНЯТИЕ О ФОРМАХ РАСПРЕДЕЛЕНИЯ

Обобщающие характеристики (показатели) центра распределения и степени вариации не дают представления о форме распределения, так как не вскрывают характера изменения частот. Для выражения особенностей формы распределения применяются показатели асимметрии и эксцесса, кривые распределения.

Для сравнительного анализа степени асимметрии нескольких распределений рассчитывается относительный показатель асимметрии (A_S):

$$A_S = \frac{\bar{x} - Mo}{\sigma}.$$

Величина показателя асимметрии A_S может быть положительной и отрицательной. Положительная величина показателя асимметрии указывает на наличие правосторонней асимметрии. Отрицательный знак показателя асимметрии говорит о наличии лево-

сторонней асимметрии. Чем больше абсолютная величина коэффициента, тем больше степень скошенности. Принято считать, что если коэффициент асимметрии меньше 0,25, то асимметрия незначительная, если свыше 0,5, то асимметрия значительная.

Наиболее распространенным является показатель асимметрии, исчисляемый по формуле:

$$A_S = \frac{\mu_3}{\sigma^3}.$$

где μ_3 — центральный момент третьего порядка;

$$\mu_3 = \frac{\sum (x - \bar{x})^3 \times f}{\sum f}.$$

Показатель асимметрии не только определяет степень асимметрии, но и указывает на наличие или отсутствие асимметрии в распределении признака в генеральной совокупности. Оценка степени существенности этого показателя дается с помощью средней квадратической ошибки, рассчитываемой по формуле:

$$\sigma_{A_S} = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}}.$$

где n — число наблюдений.

Если $\frac{|A_S|}{|\sigma_{A_S}|} > 3$, то асимметрия существенна и распределение признака в генеральной совокупности не является симметричным.

Если $\frac{|A_S|}{|\sigma_{A_S}|} < 3$, то асимметрия несущественна, и ее наличие объясняется влиянием случайных обстоятельств.

Для симметричных распределений рассчитывается показатель эксцесса (островершинности):

$$E_x = \frac{\mu_4}{\sigma^4} - 3.$$

где μ_4 — центральный момент четвертого порядка;

$$\mu_4 = \frac{\sum (x - \bar{x})^4 \times f}{\sum f}.$$

Эксцесс может быть положительным и отрицательным. У высоковершинных распределений показатель эксцесса имеет положительный знак (+), а у низкововершинных — отрицательный знак (-). Предельным значением отрицательного эксцесса является значение $E_x = -2$. Величина положительного эксцесса является бесконечной величиной. В нормальном распределении $\mu_4/\sigma^4 = 3$. Следовательно, для нормального закона $E_x = 0$.

Средняя квадратическая ошибка эксцесса исчисляется по формуле:

$$\sigma_{E_x} = \sqrt{\frac{24n(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}}.$$

где n — число наблюдений.

Все перечисленные характеристики играют большую роль при анализе вариационных рядов и определении типа кривой распределения, а также при выравнивании вариационных рядов.

§ 21.7. КРИВЫЕ РАСПРЕДЕЛЕНИЯ

Наиболее надежный путь выявления закономерностей распределения — увеличение количества наблюдений. По мере увеличения количества наблюдений (в пределах той же однородной совокупности) при одновременном уменьшении величины интервала закономерность, характерная для данного распределения, будет выступать все более и более ясно, а представляющая полигон частот ломаная линия будет приближаться к некоторой плавной линии и в пределе должна превратиться в кривую линию.

Кривая линия, которая отражает закономерность изменения частот в чистом, исключаящем влияние случайных факторов виде, называется **кривой распределения**.

В статистической практике большой интерес представляет решение вопроса о том, в какой мере можно считать полученное в результате статистического наблюдения распределение признака в исследуемой совокупности, соответствующее нормальному распределению.

Для решения этого вопроса следует рассчитать теоретические частоты нормального распределения, то есть те частоты, которые были бы, если бы данное распределение в точности следовало закону нормального распределения. Для расчета теоретических частот применяется следующая формула:

$$f_t = \frac{ni}{\sigma} \times \frac{1}{\sqrt{2\pi}} \times \exp(-t^2/2),$$

где t — нормированное отклонение;

$$t = \frac{x - \bar{x}}{\sigma}.$$

Величина $\frac{1}{\sqrt{2\pi}} \times \exp(-t^2/2)$ определяется по специальной таблице.

Следовательно, в зависимости от величины t для каждого интервала эмпирического ряда определяются теоретические частоты.

Для проверки близости теоретического и эмпирического распределений используются специальные показатели, называемые **критериями согласия**. Наиболее распространенным является критерий согласия К. Пирсона χ^2 («хи-квадрат»), исчисляемый по формуле:

$$\chi^2 = \sum \frac{(f - f')^2}{f'}$$

где f — эмпирические частоты (частости) в интервале; f' — теоретические частоты (частости) в интервале.

Полученное значение критерия ($\chi^2_{\text{расч}}$) сравнивается с табличным значением ($\chi^2_{\text{табл}}$). Последнее определяется по специальной таблице в зависимости от принятой вероятности (P) и числа степеней свободы k (для нормального распределения k равно числу групп в ряду распределения минус 3).

Если $\chi^2_{\text{расч}} \leq \chi^2_{\text{табл}}$, то гипотеза о близости эмпирического распределения к нормальному не отвергается.

При расчете критерия Пирсона необходимо соблюдать условия: число наблюдений должно быть достаточно велико ($n \geq 50$); если теоретические частоты в некоторых интервалах меньше 5, то интервалы объединяют так, чтобы частоты были больше 5.

Используя величину χ^2 , В. И. Романовский предложил оценивать близость эмпирического распределения кривой нормального распределения по отношению:

$$\frac{\chi^2 - (m - 3)}{\sqrt{2(m - 3)}}$$

где m — число групп; $m - 3$ — число степеней свободы при исчислении частот нормального распределения.

Если $\frac{\chi^2 - (m - 3)}{\sqrt{2(m - 3)}} < 3$, то можно принять гипотезу о нормальном характере эмпирического распределения.

Распространенным критерием согласия является критерий А. Н. Колмогорова:

$$\lambda = \frac{D}{\sqrt{n}}$$

где D — максимальное значение разности между накопленными эмпирическими и теоретическими частотами; n — сумма эмпирических частот.

По таблице значений вероятностей λ -критерия находят соответствующую вероятность (P). Если найденной величине λ соответствует значительная по величине вероятность (P), то расхождения между эмпирическим и теоретическим распределениями несутельственны.

Глава 22

ВЫБОРОЧНОЕ НАБЛЮДЕНИЕ

§ 22.1. СУЩНОСТЬ ВЫБОРОЧНОГО НАБЛЮДЕНИЯ

В ряде случаев относительные и средние величины совокупности рассчитываются на основе данных выборочного наблюдения, получившего в настоящее время широкое применение в работе органов государственной статистики, научно-исследовательских институтов и предприятий.

При строгом соблюдении условий случайности и достаточно большой численности отобранных единиц выборочное наблюдение репрезентативно (представительно). По результатам изучения определенной части единиц с достаточной для практики степенью точности можно судить о всей совокупности. Однако вычисленные по материалам выборочного наблюдения статистические показатели не будут точно совпадать с соответствующими характеристиками для всей совокупности (генеральной совокупности). Величина этих отклонений называется **ошибкой наблюдения**, которая складывается из ошибок двоякого рода: ошибки регистрации (точности) и ошибки репрезентативности.

Ошибки регистрации свойственны любому наблюдению (сплошному и несплошному). Они вызваны несовершенством измерительных приборов, недостаточной квалификацией работников, неточностью подсчетов и т. п. Однако при выборочном наблюдении они значительно меньше, так как в этом случае используются более подготовленные кадры.

Ошибки репрезентативности свойственны только несплошным наблюдениям. Они характеризуют размер расхождений между величинами показателя, полученного в выборочной и генеральной совокупности в условиях одинаковой точности единичных наблюдений. Ошибки репрезентативности могут быть систематическими и случайными. Систематические ошибки возникают при нарушении установленных правил отбора единиц. Случайные ошибки репрезентативности возникают в результате недостаточно равномерного представления в выборочной совокупности различных категорий единиц генеральной совокупности.