**7.9. Проверка гипотезы о законе распределения**

Одна из важнейших задач анализа вариационных рядов заключается в выявлениизакономерности распределения и определении ее характера. Основной путь в выявлениизакономерности распределения - построение вариационных рядов для достаточно больших со-вокупностей. Большое значение для выявления закономерностей распределения имеетправильное построение самого вариационного ряда: выбор числа групп и размера интервалаварьирующего признака.

Когда мы говорим о характере, типе закономерности распределения, то имеем в виду отражениев нем общих условий, определяющих вариацию. При этом речь всегда идет о распределениях качественно однородных явлений. Общие условия, определяющие тип закономерностираспределения, познаются анализом сущности явления, тех его свойств, которые определяют вариацию изучаемого признака. Следовательно, должна быть выдвинута какая-то научнаягипотеза, обосновывающая определенный тип теоретической кривой распределения.

Под теоретической кривой распределения понимается графическое изображение ряда в виде непрерывной линии изменения частот в вариационном ряду, функционально связанного сизменением вариантов (значений признака). Теоретическое распределение может быть выражено аналитически - формулой, которая связывает частоты вариационного ряда и соответствующие значения признака. Такие алгебраические формулы носят название законов распределения.

Большое познавательное значение имеет сопоставление фактических кривых распределения с теоретическими.

Как уже отмечалось, часто пользуются типом распределения, которое называется нормальным.Формула функции плотности нормального распределения:

http://www.stathelp.ru/ots/img377.jpg.

Следовательно, кривая нормального распределения может быть построена по двум параметрам- средней арифметической ц и среднему квадратическому отклонению ст.

Гипотезы о распределениях заключаются в том, что выдвигается предположение о том, чтораспределение в генеральной совокупности подчиняется какому-то определенному закону.Проверка гипотезы состоит в том, чтобы на основании сравнения фактических (эмпирических)частот с предполагаемыми (теоретическими) частотами сделать вывод о соответствиифактического распределения гипотетическому распределению. Может проводиться и сравнениечастостей.

Под гипотетическим распределением необязательно понимается нормальное распределение.Может быть выдвинута гипотеза о биномиальном распределении, распределении Пуассона и т.д.Причина частого обращения к нормальному распределению в том, что в этом типе распределения выражается закономерность, возникающая при взаимодействии множестваслучайных причин, когда ни одна из них не имеет преобладающего влияния. Закон нормальногораспределения лежит в основе многих теорем математической статистики, применяемых дляоценки репрезентативности выборок, при измерении связей и т. д. В социально-экономическойстатистике нормальное распределение встречается редко, но сравнение с ним важно длявыяснения степени и характера отклонения от него фактического распределения.

В главе 5 отмечалось, что близость средней арифметической величины, медианы и модыуказывает на вероятное соответствие изучаемого распределения нормальному закону. Но болееполная и точная проверка соответствия распределения гипотезе о нормальном законепроизводится с использованием специальных критериев, из которых рассмотрим наиболееупотребимый критерий χ2 (хи-квадрат) К. Пирсона.

Для проверки гипотезы о соответствии эмпирического распределения закону нормальногораспределения необходимо частоты (частости) фактического распределения сравнить счастотами (частостями) нормального распределения. Значит, нужно по фактическим даннымвычислить теоретические частоты кривой нормального распределения f? по формуле (длядискретных рядов):

http://www.stathelp.ru/ots/img378.jpg, (7.27)

где п - объем выборки;

i - величина интервала вариационного ряда.

Значение ординат кривой нормального распределения f(t) можно получить по таблицамзначения функции:

http://www.stathelp.ru/ots/img379.jpg.

Проверяемая гипотеза формулируется как Н0: fj = f?j альтернаивная - как Н1: fj ? f?j.

Проверка гипотезы требует, чтобы был построен теоретический ряд распределения с частотамиf?j, соответствующими нормальному закону, при тех же значениях параметров распределения

http://www.stathelp.ru/ots/img380.jpg

Методика построения теоретического ряда такова:

1. По фактическому интервальному ряду (табл. 5.6) вычисляются значения / для каждойгруппь< хозяйств по формуле (для интервальных рядов):

http://www.stathelp.ru/ots/img381.jpg -для начала и конца интервала.

2. Вычисляется вероятность попадания единицы наблюдения в данный интервал привыполнении гипотезы о нормальном законе:

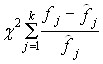
http://www.stathelp.ru/ots/img382.jpg,

где |tj| > |tj+1|

3. Определяется теоретическая частота в данной группе, равная произведению объемасовокупности на вероятность попадания в данный интервал:

http://www.stathelp.ru/ots/img383.jpg

4. Находится значение критерия χ2 по формуле

 (7.28)

где k — число категорий ряда распределения;

j - номер категории;

fj - частота эмпирического распределения;

f?j - частота теоретического распределения.

При расчете χ2 частоты можно заменить частостями:

http://www.stathelp.ru/ots/img385.jpg (7.29)

где pj - частости эмпирического распределения;

πj - вероятности теоретического распределения.

При этом, согласно Ф. Йейтсу (Jates), группы с теоретическими частотами менее 5 принятообъединять, что снижает влияние случайных ошибок (см. [6]).

Если все эмпирические частоты равны соответствующим теоретическим частотам, то χ2 равнонулю. Очевидно, что чем больше отличаются эмпирические и теоретические частоты, тем χ2больше; если расхождение несущественно, то χ2 должно быть малым. Имеются специальныетаблицы критических значений χ2 при 5%-ном и 1%-ном уровнях значимости. Критическиезначения зависят от числа степеней свободы (d.f. - degrees of freedom) и уровня значимости.

Число степеней свободы рассчитывается так: если эмпирический ряд распределения имеет kкатегорий, то k эмпирических частот f1, f2, …, fk должны быть связаны следующимсоотношением: http://www.stathelp.ru/ots/img386.jpg Если параметры теоретического распределения известны, тотолько k - 1 частот могут принимать произвольные значения, т. е. свободно варьировать, а последняя частота может быть найдена из указанного соотношения. Поэтому говорят, что система из k частот благодаря наличию одной связи теряет одну «степень свободы» и имеет только k — 1 степеней свободы. Кроме того, если при нахождении теоретических частот рпараметров теоретического распределения неизвестны, то они должны быть найдены по даннымэмпирического ряда. Это накладывает на эмпирические частоты еще р связей, благодаря чемусистема теряет еще р степеней свободы. Таким образом, число свободно варьируемых частот (азначит, и число степеней свободы) становится равным:

d.f. = (k - 1) - р = k - (р + 1). (7.30)

Полученное значение критерия χ2 сравнивается с табличным при числе степеней свободы,равном числу групп (с условием Ф. Йейтса), за минусом трех - по числу фиксированныхпараметров в формуле нормального закона распределения и с учетом равенства суммтеоретических и фактических частот (см. приложение, табл. 4).

В первой графе этой таблицы дано число степеней свободы, а в заголовках граф - уровнизначимости. Если фактическое значение χ2 превышает табличное при том же числе степенейсвободы, то вероятность соответствия распределения нормальному закону меньше указанной.Результаты расчета χ2 по данным табл. 5.6 (глава 5) приведены в табл. 7.5 при х = 30,3; s = 8,44.

Сумма теоретических частот нормального распределения меньше суммы фактических частот, таккак нормальный закон не ограничен рамками фактических минимума и максимума.

Число групп после объединения малочисленных составило 7. Критическое значение χ2 по табл. 4 приложения при 7-3 = 4 степеням свободы и значимости 0,05 составляет 9,49. Значит,вероятность расхождения распределения с нормальным меньше 0,05, и вероятностьсоответствия его нормальному закону больше 0,95. Табличное значение χ2 для значимости 0,1равно 7,78, что также больше фактического.

Таблица 7.5

Проверка соответствия распределения хозяйств по урожайности

зерновых культур нормальному закону

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Группы  хозяйств | fj | tj | tj + i | Рj | f?j | (fj - f?j)2/ f??2j |
| 1 | 6 | -2,41 | -1,81 | 0,0235 | 3 | 0,071 |
| 2 | 9 | -1,81 | -1,22 | 0,0798 | 11 |  |
| 3 | 20 - | -1,22 | -0,63 | 0,1531 | 22 | 0,182 |
| 4 | 41 | -0,63 | -0,04 | 0,2197 | 32 | 2,531 |
| 5 | 26 | -0,04 | 0,56 | 0,2282 | 33 | 1,485 |
| 6 | 21 | 0,56 | 1,15 | 0,1627 | 23 | 0,174 |
| 7 | 14 | 1,15 | 1,74 | 0,0842 | 12 | 0,333 |
| 8 | 5 | 1,74 | 2,33 | 0,0310 | 4 | 0,200 |
| 9 | 1 | 2,33 | 2,93 | 0,0082 | 1 |  |
| Σ | 143 | × | × | 0,9904 | 141 | 4,976 |

Ясно, что гипотеза о соответствии распределения хозяйств по урожайности нормальному законуне может быть отклонена.

Какое практическое значение может иметь произведенная проверка гипотезы? Во-первых,соответствие нормальному закону позволяет прогнозировать, какое число хозяйств (или долясовокупности) попадает в тот или иной интервал значений признака. Во-вторых, нормальноераспределение возникает при действии на вариацию изучаемого показателя множестванезависимых факторов. Из этого следует, что нельзя существенно снизить вариациюурожайности, воздействуя только на один-два управляемых фактора, скажем удобрения илиэнергозатраты.

С помощью критерия χ2 можно проверять не только гипотезу о согласии эмпирического распределения с нормальным законом, но и с любым другим известным законом распределения- равномерным распределением, распределением Пуассона и т. д. Например, суд рассматривает жалобу посетителей казино на то, что, по их мнению, игральная кость, которой там пользуются, фальшива, некоторые числа очков, якобы, выпадают чаще, чем другие, и этим пользуются крупье, обирающие игроков.

Суд назначает экспертизу игральной кости: эксперт делает 600 бросков и записывает число выпавших единиц, двоек, троек и т. д.

Полученное эмпирическое распределение сравнивается с теоретическим, т. е. равномерным: в правильной кости вероятность выпадения каждого числа очков должна быть равна 1/6, при 600бросках это даст по 100 выпадений каждого числа очков. С помощью критерия χ2 проверяется нулевая гипотеза о том, что различия эмпирического и теоретического распределений случайны, т. е. не являются систематическим результатом фальсификации формы кости или положенияцентра тяжести в ней; H0 : fфакт = fтеор. Результаты испытания и расчет у приводятся в табл. 7.6.

Таблица 7.6

Результаты испытания игральной кости

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Число очков  1  2  3.  4  5  6  Итого | Количество  выпадений,  fфакт  101  86  107  94  97  117  600 | fтеор  100  100  100  100  100  100  600 | fфакт - fтеор  1  -14  7  -6  -3  17  0 | (fфакт- fтеор)2= fтеор  0,01  1,96  0.49  0,36  0.09  2,89  5,80 |

Табличное значение χ2 при уровне значимости 0,05 (это вероятность ошибочного отклонения нулевой гипотезы при условии, что она верна) и при 6-2=4 степенях свободы (фиксировано 2параметра: сумма числа бросков 600 и вероятность каждого числа очков - 1/6) составляет 9,49.Вычисленное значение χ2 =5,8, что значительно ниже табличного. Следовательно, нулевая гипотеза не отклоняется: распределение бросков по числу выпавших очков нельзя считать неравномерным . Обвинение игроков против служащих казино не подтверждено достаточнонадежно, но не доказано и то, что кость правильная. Можно назначить более дорогую экспертизу - сделать 100 000 бросков кости, но можно и согласиться, что вероятность ошибочного признания правильности кости мала - всего 5% - и отклонить обвинение.

Выбор закона распределения проводится на основе теоретического анализа. Кроме того,целесообразно руководствоваться следующей рекомендацией: выражение, определяющеефункцию плотности распределения, должно зависеть от возможно меньшего числа параметров.Например, экспоненциальное распределение зависит от одного параметра - средней величины;нормальное и логнормальное распределение - от двух параметров.

<http://www.stathelp.ru/ots/g7p9.html>