Математическая статистика

Глава 1. Анализ вариационных рядов.

§1. Предмет и задачи математической статистики.

Результаты измерений (наблюдений) называют статистическими данными. В зависимости от поставленной цели все задачи математической статистики могут быть сформулированы в различных формах, среди которых типичными являются:

Определение 2. *Вся исследуемая совокупность однородных объектов называется генеральной совокупностью.*

Определение 3. *Множество из n- объектов, отобранных случайным образом из генеральной совокупности, называется выборочной совокупностью или выборкой (n- объем выборки).*

Одним из основных способов сбора статистических данных является выборочный метод.

Определение 4. *Метод, основанный на том, что по данным обследования выборки, выделенной из данной генеральной совокупности, делается заключение обо всей генеральной совокупности, называется выборочным методом.*

Определение 5. *Выборка называется репрезентативной, если каждый объект генеральной совокупности имеет одинаковую возможность попасть в выборку.*

В реальных социально - экономических системах нельзя проводить эксперименты, поэтому данные обычно представляют собой пассивные наблюдения за происходящим процессом, например: курс валюты на бирже в течение месяца, урожайность пшеницы в хозяйстве за 30 лет, производительность труда рабочих за смену и т.д.

В результате наблюдений мы получаем сведения о численной величине изучаемого признака у каждого члена данной совокупности.

§2. Вариационные ряды.

Определение 1. *Значение случайной величины, соответствующее отдельной группе сгруппированного ряда наблюдаемых данных, называется* ***вариантом*** *(* x_i *) , а изменения этого значения –* ***варьированием.***

Результаты наблюдений, в общем случае - ряд чисел, расположены в беспорядке, поэтому их необходимо упорядочить.

Определение 2. ***Вариационным рядом*** *называется ранжирование в порядке возрастания вариант с соответствующими им частотами (ранжир - в переводе с фр.- «ставить в ряд по росту»).*

Определение 3. *Операция, заключающаяся в том, что результаты наблюдений над случайной величиной располагают в порядке неубывания, называется* ***ранжированием опытных данных.***

Для каждой группы сгруппированного ряда данных можно подсчитать их численность, т.е. определить число, которое показывает, сколько раз встречается соответствующий вариант в ряде наблюдений.

Определение 4. *Численность отдельной группы сгруппированного ряда наблюдаемых данных называется* ***частотой*** *или* ***весом*** *соответствующего варианта и обозначается* m_i, *где* i *- индекс варианта.*

Определение 5. *Отношение частоты данного варианта к объему совокупности называется* ***относительной частотой*** $\hat{p}_i$ *или* ***частостью*** *этого варианта.*

$$\hat{p}_i = \frac{m_i}{n}$$

Пример 1: *Пусть мы интересуемся размерами проданной в магазине мужской обуви за некоторый отрезок времени. Получены данные в порядке продажи:*

41	39	40	38	43	41	38	41	42	40	42	41	40	42	39	41	41
36	43	42	41	38	41	40	42	41	42	42	42	40	41	41	39	42
40	40	39	41	39	38	40	41	41	40	40	40	39	42	43	37	40
42	43	42	38	40	40	41	41	41	40	43	42	42	39	43	41	40
41	42	42	43	39	41	42	43	41	42	40						

Интересующий нас признак принимает различные и притом только целые значения, причем он постоянно меняется, как говорят, варьирует.

Упорядочим записанный ряд:

1	2	3–7	8–15	16–32	33–53	54–71	72–79
36	37	38...38	39...39	40...40	41...41	42...42	43...43
		5	8	17	21	18	8

Данные о количествах и размерах проданной мужской обуви будут более наглядными, если их представить в виде таблицы.

a) b) *Таблица 1*

Размер обуви (варианты) x_i	Число проданных пар (частота) m_i		Доля покупок (частость) $\hat{p}_i$	
	частота m_i	накопленная частота S_i	частость $\hat{p}_i$	накопленная частость
36	1	1	0,013	0,013
37	1	2	0,013	0,026
38	5	7	0,063	0,089
39	8	15	0,101	0,190
40	17	32	0,215	0,405
41	21	53	0,266	0,671
42	18	71	0,228	0,899
43	8	79	0,101	1,0
Всего	n = 79	-	1,0	-

Получен вариационный ряд. Он может быть записан с указанием числа проданных пар (частот каждого варианта) (а) или указанием доли каждого из них во всей совокупности (частостей) (б).

Рассмотренный нами вариационный ряд называется дискретным.

Определение 6. ***Дискретным*** ***вариационным рядом распределения называется ранжированная совокупность вариантов с соответствующими им частотами*** m_i ***или частостями*** $\widehat{p}_i$ **.**

В общем виде его можно записать так:

x_i	x_1	x_2	…	x_n
m_i	m_1	m_2	…	m_n

Вариационный ряд часто дополнительно характеризуется накопленными частотами или накопленными частостями (таблица 1).

Определение 7. ***Накопленные частоты*** ***характеризуют число членов данной совокупности, у которых рассматриваемый признак принимает значения, не превышающие данного варианта.***

Определение 8. ***Накопленные частости*** ***– результаты последовательного суммирования частостей всех вариантов, включая частость данного варианта. Накопленная частость показывает долю членов совокупности, у которых интересующий нас признак не превосходит данного значения.***

Кроме дискретных вариационных рядов широкое применение имеют ***непрерывные (интервальные)*** ***вариационные ряды.***

Определение 9. ***Интервальным*** ***вариационным рядом называется упорядоченная совокупность интервалов варьирования значений случайной величины с соответствующими частотами или частостями попаданий в каждый из них значений случайной величины.***

Интервальный ряд целесообразно построить, если число возможных значений дискретной величины велико или признак является непрерывным, т.е. может принимать любые значения в пределах некоторого интервала.

Для построения интервального ряда необходимо определить величину частичных интервалов, на которые разбивается весь интервал варьирования наблюдаемых значений случайной величины.

Считая, что все частичные интервалы имеют одну и ту же длину, для каждого интервала следует установить его верхнюю и нижнюю границы, а затем в соответствии с полученной упорядоченной совокупностью частичных интервалов сгруппировать результаты наблюдений. Т.е. промежуток изменения признака разбивается на ряд отдельных интервалов и подсчитывается количество значений величины в каждом из них.

***Размах варьирования** определяется по формуле:*

$$W = X_{\max} - X_{\min}$$

Для определения величины частичного интервала воспользуемся формулой Стерджесса:

$h = \frac{W}{k}$ *(*), где k- число интервалов*

$k = 1 + 3{,}322 \lg n \qquad k \in [6;12]$

W-размах варьирования

Тогда формулу (*) можно записать:

$$h = \frac{X_{\max} - X_{\min}}{1 + 3{,}322 \lg n}$$

Если окажется, что ***h*** *- дробное число, то за длину частичного интервала следует брать, либо ближайшее целое число, либо ближайшую простую дробь.*

За начало первого интервала рекомендуется брать величину: $X_{нач.} = X_{\min} - \frac{h}{2}$

Конец последнего интервала $X_{кон}$ *должен удовлетворять условию:*

$$X_{кон.} - h \le X_{\max} < X_{кон.}$$

Промежуточные интервалы получают, прибавляя к концу предыдущего интервала длину частичного интервала ***h****.*

Теперь, просматривая, результаты наблюдений, определяем, сколько значений признака попало в каждый конкретный интервал. При этом в интервал включают значение случайной величины, большие или равные нижней границе и меньшие верхней

Пример 2: ***Пусть дан ряд распределения хозяйств по количеству рабочих на 100 га с/х угодий (n=60) :***

12	6	8	6	10	11	7	10	12	8	7	7	6	7	8	6	11	9	11	9	10
11	9	10	7	8	8	8	11	9	8	7	5	9	7	7	14	11	9	8	7	4
7	5	5	10	7	7	5	8	10	10	15	10	10	13	12	11	15	6			

Построить интервальный вариационный ряд.

Решение. ***Для определения числа групп подставим значение n=60 в формулу Стерджесса:***

$$k = 1 + 3,322 \lg 60 \approx 6,907; \quad k = 7$$

Найдем длину частичного интервала

$$h = \frac{X_{\max} - X_{\min}}{k} = \frac{15-4}{7} = \frac{11}{7} \approx 1,6$$

Построим интервальный вариационный ряд, для этого в качестве начального значения используем X_{min}

Группы хозяйств по численности работников на 100 га с/х угодий		*Число хоз-в в группе* m_i	*Накопленное число хоз-в* S_i	*Относительная частота* $\hat{p}_i$
4- 5,6		*5*	*5*	*5/60*
5,61- 7,2		*17*	*22*	*17/60*
7,21- 8,8		*9*	*31*	*9/60*
8,81- 10,4		*15*	*46*	*15/60*
10,41- 12,0		*10*	*56*	*10/60*
12,01- 13,6		*1*	*57*	*1/60*
13,61- 15,2		*3*	*60*	*3/60*
Итого		*60*	*-*	*1*

Иногда интервальный вариационный ряд для простоты исследований условно заменяют дискретным.

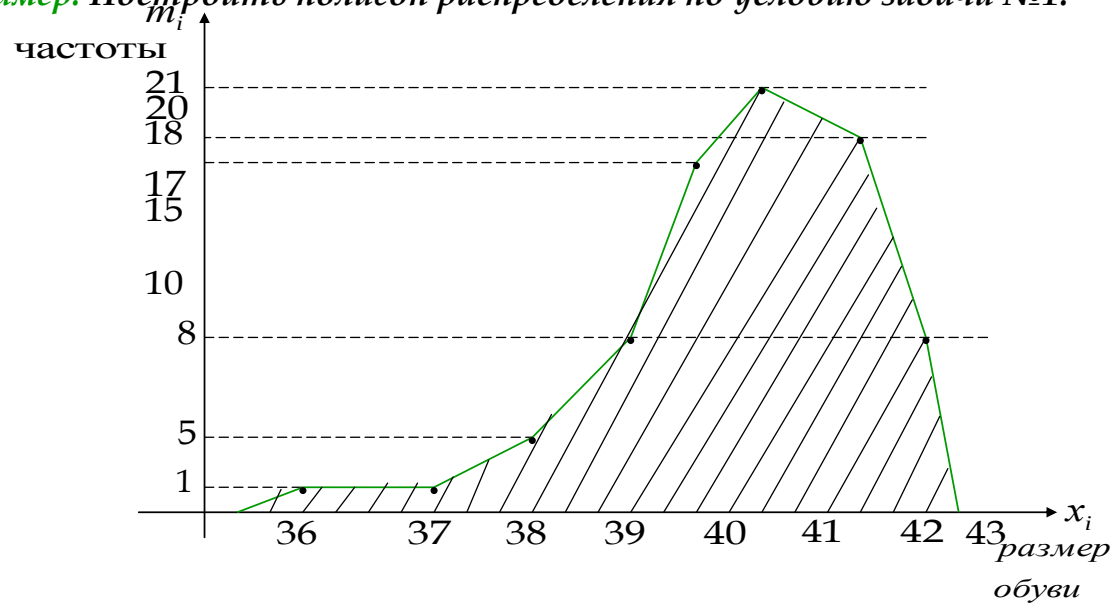
В этом случае серединное значение i -го интервала принимают за вариант x_i , а соответствующую интервальную частоту m_i - за частоту этого интервала.

§3. Графическое изображение вариационных рядов.

Графическое изображение позволяет представить в наглядной форме закономерности варьирования значений признаков с помощью полигона, гистограммы, кумуляты и огивы.

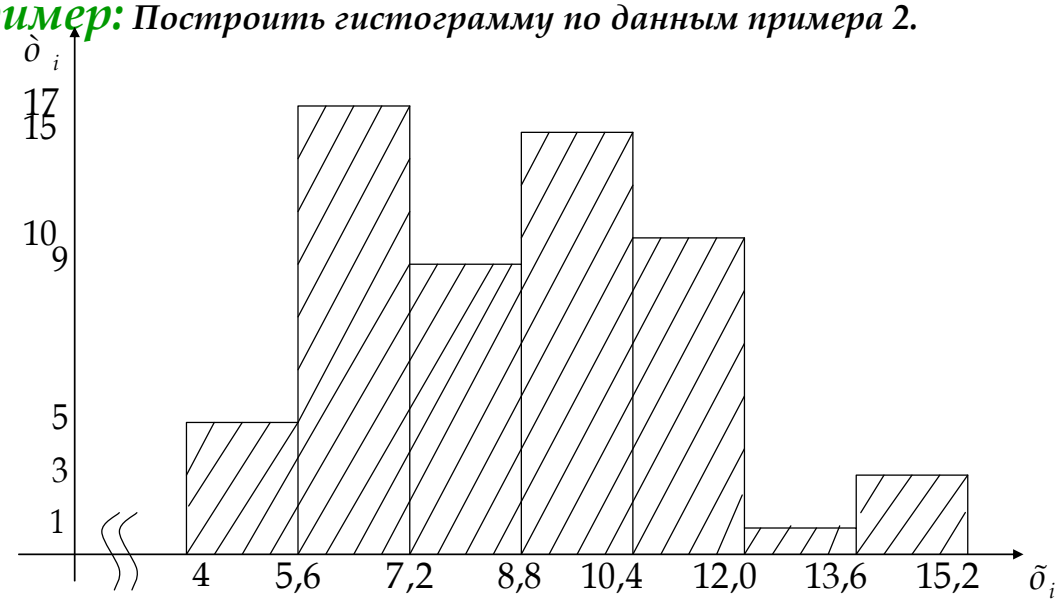
Определение 1. ***Полигоном*** *(для дискретного вариационного ряда) называется ломанная, соединяющая на плоскости точки с координатами $(x_i; m_i)$.*

Пример: *Построить полигон распределения по условию задачи №1.*



Определение 2. ***Гистограммой*** *(для интервального вариационного ряда) называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат интервалы* $(x_{i-1}; x_i)$ *, а высотами - частоты* m_i *.*

Пример: *Построить гистограмму по данным примера 2.*

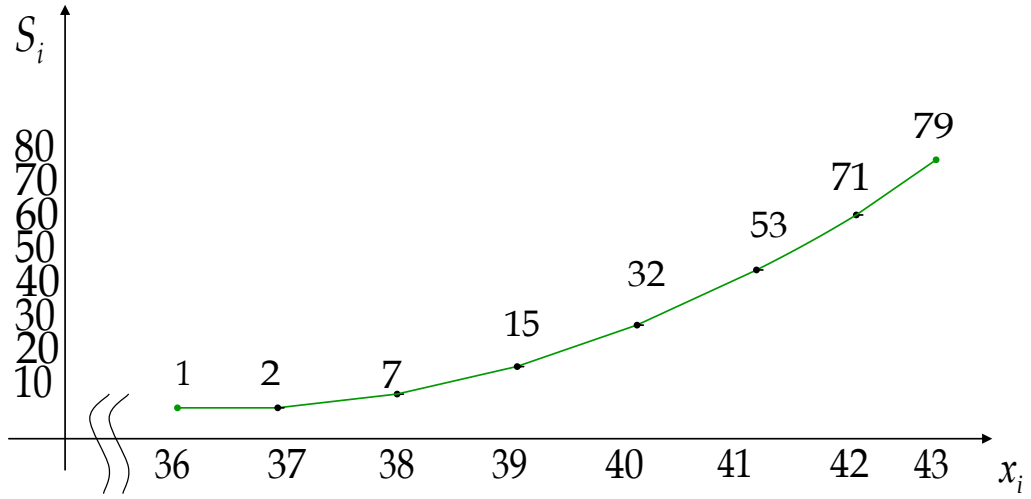


Если в вариационном ряду вместо частот взяты соответственно накопленные частоты, то полученный ряд называется ***кумулятивным рядом*** *(кумуляция – от латинского «скопление»).*

Определение 3. ***Кумулятой*** *называется ломанная, соединяющая на плоскости точки вида* (x_i, S_i)*.*

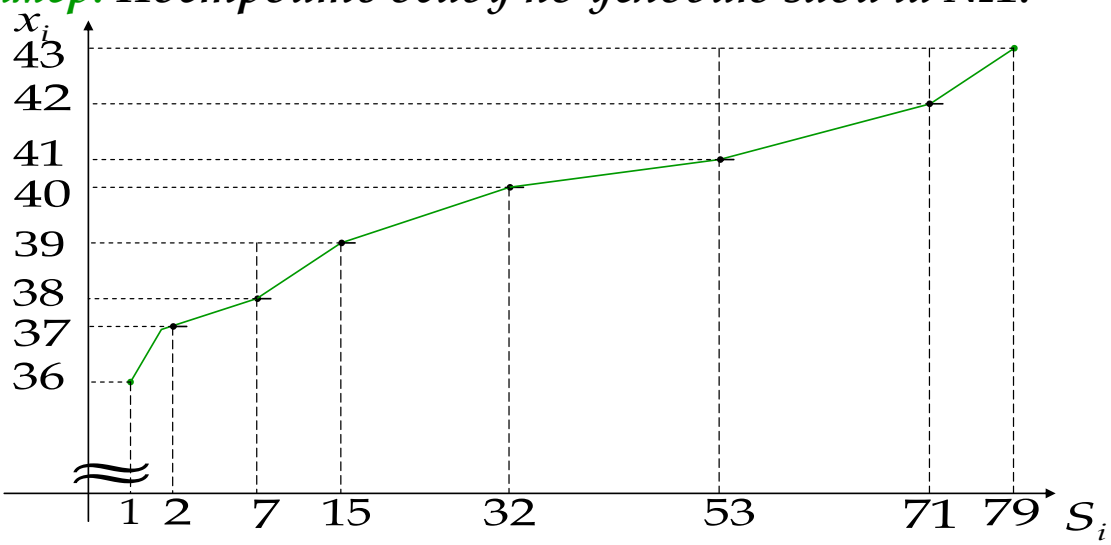
Кумуляту иначе называют ***полигоном накопленных частот*****.**

Пример: ***Построить кумулятивную кривую по условию задачи №1.***



Определение 4. ***Если по оси абсцисс откладывать накопленные частоты, а по оси ординат - значение признака, затем полученные точки соединить отрезками, то получится огива.***

Пример: ***Построить огиву по условию задачи №1.***



§4. Числовые характеристики вариационных рядов.

Вариационные ряды позволяют получить первое представление об изучаемом распределении. Далее необходимо исследовать числовые характеристики распределения (аналогичные характеристикам распределения теории вероятностей): характеристики положения (средняя арифметическая, мода, медиана); характеристики рассеивания (дисперсия, среднее квадратическое отклонение, коэффициент вариации); характеристики меры скошенности (коэффициент асимметрии) и островершинности (эксцесс) распределения.

Определение 1. ***Средней арифметической*** $\left(\overline{X}\right)$ ***дискретного вариационного ряда называется отношение суммы произведений вариантов на соответствующие частоты к объему совокупности:***

$$\overline{X} = \frac{\sum x_i m_i}{n} \qquad (1)$$

Пример: ***Найти*** $\left(\overline{X}\right)$ ***по условию задачи 1.***

$$\overline{X} = \frac{36\cdot 1 + 37\cdot 1 + 38\cdot 5 + 39\cdot 8 + 40\cdot 17 + 41\cdot 21 + 42\cdot 18 + 43\cdot 8}{79} = \frac{3206}{79} = 40,5$$

Вычисленное по формуле (1) среднее арифметическое называется взвешенным, так как частоты m_i ***называются весами, а операция умножения*** x_i ***на*** m_i ***- взвешиванием.***

Для интервального вариационного ряда за x_i ***принимают середину i-го интервала, а за*** m_i ***– соответствующую интервальную частоту.***

Пример: ***Найти*** $\left(\overline{X}\right)$ ***по условию задачи 2.***

$$\overline{X} = \frac{4{,}8\cdot 5 + 6{,}4\cdot 17 + 8{,}0\cdot 9 + 9{,}6\cdot 15 + 11{,}2\cdot 10 + 12.8\cdot 1 + 14{,}4\cdot 3}{60} = \frac{516{,}8}{60} \approx 8{,}613$$

Определение 2. ***Модой*** $\left(\widehat{M}_0(x)\right)$***дискретного вариационного ряда называется вариант, имеющий наибольшую частоту.***

Пример: ***Найти*** $\widehat{M}_0(x)$ ***по условию задачи 1.***

$m_{\max} = 21$ соответствует $x = 41 \Rightarrow \widehat{M}_0(x) = 41$

Для интервальных вариационных рядов при нахождении $\widehat{M}_0(x)$ ***используют формулу:***

$$\widehat{M}_0(x) = x_0 + h \cdot \frac{m_i - m_{i-1}}{\left(m_i - m_{i-1}\right) + \left(m_i - m_{i+1}\right)}$$

,где

x_0 ***- начало модального интервала;***

h ***- длина частичного интервала;***

m_i ***- частота модального интервала;***

m_{i-1} ***- частота предмодального интервала;***

m_{i+1} ***- частота послемодального интервала.***

Пример: ***Найти*** $\hat{M}_0(x)$ ***по условию задачи 2.***

$m_{\max} = 17$ соответствует интервалу $5,61 - 7,2$

$$\hat{M}_0(x) = 5,61 + 1,6 \cdot \frac{17-5}{(17-5)+(17-9)} \approx 6,56$$

Определение 3. ***Медианой*** $\left(\hat{M}_e(x)\right)$ ***дискретного вариационного ряда называется вариант, делящий ряд на две равные части.***

Если дискретный вариационный ряд имеет четное (2n) число членов, то:

$$\hat{M}_e(x) = \frac{x_n + x_{n+1}}{2}$$

Если дискретный вариационный ряд имеет нечетное (2n-1) число значений варьирующего признака, расположенных в порядке возрастания, то медианой этого распределения является вариант x_n

$$\hat{M}_e(x) = x_n$$

Математическая статистика

Глава 1. Анализ вариационных рядов.

§4. Числовые характеристики вариационных рядов.

Пример: **Найти** $\widehat{M}_e(x)$ **по условию задачи 1.**

$$n = 79 \quad 2n - 1 = 79 \quad \Rightarrow \quad 2n = 80 \quad \Rightarrow \quad n = 40$$
$$x_{40} = 41 \quad \Rightarrow \quad \widehat{M}_e(x) = 41$$

При нахождении $\widehat{M}_e(x)$ ***для интервальных вариационных рядов используют формулу:***

$$\widehat{M}_e(x) = x_0 + h \cdot \frac{0,5n - S_{i-1}}{m_i},$$

где

x_0 - *начало медианного интервала;*

h - *длина частичного интервала;*

n - *объем совокупности;*

S_{i-1} - *накопленная частота интервала, предшествующего медианному;*

m_i - *частота медианного интервала.*

Пример: ***Найти*** $\widehat{M}_e(x)$ ***по условию задачи 2.***

$n=60 \Rightarrow \frac{n}{2}=30 \Rightarrow$ ***медиана расположена в интервале*** $7,21-8,8$

$$\widehat{M}_e(x)=7,21+1,6\frac{0,5\cdot 60-22}{9}\approx 8,62$$

Определение 4. ***Дисперсия*** ***вариационного ряда (как дискретного, так и интервального) характеризует средний квадрат отклонения значения признака от его среднего значения.***

$$D(x)=\frac{\sum\left(x_i-\bar{x}\right)^2\cdot m_i}{n}$$

Определение 5. ***Среднее квадратическое отклонение*** ***вариационного ряда распределения характеризует те же значения, что и дисперсия, но измеряется в единицах варьирующего признака.***

$$\sigma(x)=\sqrt{\frac{\sum\left(x_i-\bar{x}\right)^2\cdot m_i}{n}}$$

Определение 6. ***Коэффициент вариации*** *характеризует относительное значение среднего квадратического отклонения и служит для сравнения колеблемости несоизмеримых показателей.*

$$V = \frac{\sigma(x)}{\overline{X}} \cdot 100\%$$

Моменты *для вариационных рядов в математической статистике находятся по формулам, аналогичным формулам из теории вероятностей:*

$$\hat{\nu}_k = \frac{\sum x_i^k \cdot m_i}{n} = \sum x_i^k \cdot \hat{p}_i$$

- начальный момент k- го порядка.

$$\hat{M}_k = \frac{\sum \left(x_i - \overline{x}\right)^k \cdot m_i}{n}$$

- центральный момент k-го порядка.

Соотношения между начальными и центральными моментами:

Коэффициент асимметрии – $$\hat{A} = \frac{\sum \left(x_i - \overline{x}\right)^3 \cdot m_i}{n \cdot \sigma^3(x)}$$

Эксцесс - $$\hat{E} = \frac{\sum \left(x_i - \overline{x}\right)^4 \cdot m_i}{n \cdot \sigma^4(x)} - 3$$

Пример: ***Рассчитать дисперсию, среднее квадратическое отклонение, коэффициенты вариации, асимметрии и эксцесс для задачи 2. Сделать выводы.***

Построим вспомогательную таблицу.

Группы хоз-в по численности работников на 100 га с/х угодий, чех.	**Среднее значение интервала,** x_i	**Число хоз-в в группе** m_i	$x_i m_i$	$x_i - \bar{x}$	$\left(x_i - \bar{x}\right)^2 \cdot m_i$	$\frac{x_i - \bar{x}}{\sigma(x)}$	$\left(\frac{x_i - \bar{x}}{\sigma(x)}\right)^3 \cdot m_i$	$\left(\frac{x_i - \bar{x}}{\sigma(x)}\right)^4 \cdot m_i$
4-5,6	*4,8*	*5*	*24*	*-3,813*	*72,708*	*-1,559*	*-18,954*	*29,554*
5,61-7,2	*6,4*	*17*	*108,8*	*- 2,213*	*83,280*	*- 0,905*	*-12,601*	*11,404*
7,21-8,8	*8*	*9*	*72*	*- 0,613*	*3,386*	*- 0,251*	*- 0,142*	*0,036*
8,81-10,4	*9,6*	*15*	*144*	*0,987*	*14,603*	*0,403*	*0,985*	*0,397*
10,41-12,0	*11,2*	*10*	*112*	*2,587*	*66,908*	*1,058*	*11,832*	*12,514*
12,01-13,6	*12.8*	*1*	*12,8*	*4,187*	*17,528*	*1,712*	*5,017*	*8,588*
13,61-15,2	*14,4*	*3*	*43,2*	*5,787*	*100,457*	*2,366*	*39,740*	*94,030*
Итого	*-*	*60*	*516,8*	*0*	*358,869*	*-*	*25,876*	*156,523*

$$\bar{x} = \frac{516,8}{60} = 8,613$$

$$D(x) = \frac{\sum \left(x_i - \bar{x}\right)^2 \cdot m_i}{n} = \frac{358,869}{60} = 5,981$$

$$\sigma(x) = \sqrt{D(x)} = \sqrt{5,981} \approx 2,446$$

$$V = \frac{\sigma(x)}{\bar{x}} \cdot 100\% = \frac{2,446}{8,613} \cdot 100\% = 28,4\%$$

Таким образом, средняя численность работников на 100 га с/х угодий по исследуемой совокупности хозяйств составила 8,61 чел. Плотность работников в среднем колебалась в промежутке $\boxed{\bar{x} \pm \sigma(x)}$ ***= 8,61±2,45, т.е. от 6,16 до 11,06 чел. на 100 га с/х угодий.***

Этот интервал, а так же коэффициент вариации показывает, что имеются большие различия в обеспечении хозяйств рабочей силой.

$$\widehat{A} = \frac{\sum \left(x_i - \bar{x} \right)^3 \cdot m_i}{n \cdot \sigma^3 \left(x \right)} = \frac{25,876}{60} = 0,43$$

$$\widehat{E} = \frac{\sum \left(x_i - \bar{x} \right)^4 \cdot m_i}{n \cdot \sigma^4 \left(x \right)} - 3 = \frac{156,523}{60} - 3 = -0,39$$

Найденное значение коэффициента асимметрии (не достаточно близкое к нулю) указывает, что распределение не симметрично. Эксцесс также отличен от нуля, что говорит о возможном отличии распределения от нормального.

§5. *Выборочный метод.*

В реальных условиях обычно бывает трудно или экономически нецелесообразно, а иногда и невозможно, исследовать всю совокупность, характеризующую изучаемый признак (генеральную совокупность). Поэтому на практике широко применяется выборочное наблюдение, когда обрабатывается часть генеральной совокупности (выборочная совокупность).

Свойства (закон распределения и его параметры) генеральной совокупности неизвестны, поэтому возникает задача их оценки по выборке. Для получения хороших оценок характеристик генеральной совокупности необходимо, чтобы выборка была репрезентативной (представительной). Репрезентативность в силу закона больших чисел, достигается случайностью отбора.

Различают 5 основных типов выборок:

1. Собственно - случайная:

- ***а) повторная (элементы после выбора возвращаются обратно);***
- ***б) бесповторная (выбранные элементы не возвращаются).***

2. *Типическая – генеральная совокупность предварительно разбивается на группы типических элементов, и выборка осуществляется из каждой.*

   Следует различать:

   а) равномерные выборки (при равенстве объемов исходных групп в генеральной совокупности выбирается одинаковое количество элементов из каждой);

   б) пропорциональные (численность выборок формируют пропорционально численностям или средним квадратическим отклонениям групп генеральной совокупности);

   в) комбинированные (численность выборок пропорциональна и средним квадратическим отклонениям, и численностям групп генеральной совокупности).

3. *Механическая – отбор элементов проводится через определенный интервал.*
4. *Серийная – отбор проводится не по одному элементу, а сериями для проведения сплошного обследования.*
5. *Комбинированная – используются различные комбинации вышеуказанных методов, например, типическая выборка сочетается с механической и собственно случайной.*

После осуществления выборки возникает задача оценки числовых характеристик генеральной совокупности по элементам выборочной совокупности. Различают точечные и интервальные оценки.

Определение 1. ***Точечной оценкой*** *характеристики генеральной совокупности называется число, определяемое по выборке.*

Пусть $\widehat{\Theta} = \widehat{\Theta}_n$ *выборочная характеристика, вычисленная по результатам* n *наблюдений величины* X*, используемая в качестве оценки* Θ *- характеристики генеральной совокупности (в качестве* Θ *может быть* $M(X); D(X)$ *и т.д.).*

Качество оценки $\widehat{\Theta}$ *устанавливается по 3-м свойствам:*

1) ***Состоятельность.*** *Оценка* $\widehat{\Theta}_n$ *является состоятельной оценкой генеральной совокупности* Θ*, если для любого* $\varepsilon > 0$ *выполняется неравенство:*

$$\lim_{n\to\infty} P\left(\left|\widehat{\Theta}_n - \Theta\right| < \varepsilon\right) = 1$$

Это означает, что при увеличении объема выборки n *выборочная характеристика стремится к соответствующей характеристике генеральной совокупности* $\left(\widehat{\Theta}_n \to \Theta\right)$

2) ***Несмещенность.*** ***Оценка*** $\widehat{\Theta}_n$ ***генеральной характеристики*** Θ ***называется несмещенной, если для любого фиксированного числа наблюдений*** n ***выполняется равенство:***

$$M\left(\widehat{\Theta}_n\right) = \Theta$$

3) ***Эффективность.*** ***Несмещенная оценка*** $\widehat{\Theta}_n$ ***генеральной характеристики*** Θ ***называется несмещенной эффективной, если среди всех подобных оценок той же характеристики она имеет наименьшую дисперсию:***

$$D\left(\widehat{\Theta}_n\right) \to \min$$

Статистики $\overline{x}$ ***и*** $\hat{p}_i$ ***являются состоятельными, несмещенными и эффективными характеристиками математического ожидания*** $M(X)$ ***и вероятности*** P ***соответственно.***

Выборочная дисперсия $D(x) = \sigma^2(x)$ ***не обладает свойством несмещенности.***

На практике используют исправленную выборочную дисперсию S^2***, которая является несмещенной оценкой дисперсии генеральной совокупности:***

$$S^2 = \frac{n}{n-1} \cdot \sigma^2(x) = \frac{n}{n-1} \cdot \frac{\sum\left(x_i - \overline{x}\right)^2 \cdot m_i}{n} = \frac{\sum\left(x_i - \overline{x}\right)^2 \cdot m_i}{n-1} \Rightarrow$$

$$S^2 = \frac{\sum\left(x_i - \overline{x}\right)^2 \cdot m_i}{n-1}$$ ***, где***

S– стандартное отклонение.

Кроме того, в расчетах используют ***стандартную ошибку выборки:***

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

Определение 2. ***Интервальной*** *называют оценку, которая определяется двумя числами - границами интервала.*

Интервальная оценка позволяет ответить на вопрос: внутри какого интервала, и с какой вероятностью находится неизвестное значение оцениваемого параметра генеральной совокупности?

Пусть $\tilde{\Theta}$ *точечная оценка параметра* Θ*. Чем меньше разность* $\tilde{\Theta} - \Theta$ *, тем точнее и лучше оценка. Обычно говорят о* ***доверительной вероятности*** $p=1-\alpha$ *, с которой* Θ *будет находиться в интервале*

$$\tilde{\Theta} - \Delta < \Theta < \tilde{\Theta} + \Delta$$ *, где*

$\Delta(\Delta>0)$ *- предельная ошибка выборки, которая может быть задана наперед, либо вычислена;*

α *- риск или уровень значимости (вероятность того, что неравенство будет неверным).*

В качестве $(1-\alpha)$ *принимают значения 0,90;0,95; 0,99;0,999. Доверительная вероятность показывает, что в* $(1-\alpha)\cdot 100\%$ *случаев оценка будет накрываться указанным интервалом.*

Точечная оценка *математического ожидания* $M(X)=a$ *определяется как средняя арифметическая*

$$\bar{x} = \frac{1}{n}\sum x_i \cdot m_i$$

Точечная оценка вероятности p_i *определяется как относительная частота:*

$$p_i = \frac{m_i}{n}$$

Для построения доверительного интервала параметра a *- математического ожидания нормального распределения, составляют выборочную характеристику (статистику), функционально зависимую от наблюдений и связанную с* a *, например:*

1. для повторного отбора:

$$u = \frac{\bar{x} - a}{\frac{\sigma(x)}{\sqrt{n}}}$$

Статистика u *распределена по нормальному закону распределения с математическим ожиданием* $a=0$ *и средним квадратическим отклонением* $\sigma(x)=1$*.*

Отсюда:

$P\left(|u| < U_{\alpha/2}\right) = 1 - \sigma(x)$ *или ,* $2\Phi\left(U_{\alpha/2}\right) = 1 - \sigma(x)$ *где*

Ф- функция Лапласа.

$U_{\alpha/2}$*- квантиль нормального закона распределения, соответствующая уровню значимости* α *.*

Доверительный интервал для параметра a*:*

$$\bar{x} - U_{\alpha/2} \cdot \frac{\sigma(x)}{\sqrt{n}} < a < \bar{x} + U_{\alpha/2} \cdot \frac{\sigma(x)}{\sqrt{n}}$$

2. Для *бесповторного отбора:*

Доверительный интервал для средней:

$$\bar{x} - \Delta_{\bar{x}} < \bar{x}_0 < \bar{x} + \Delta_{\bar{x}}$$, *где*

$\bar{x}$ - *выборочная средняя;*

$\bar{x}_0$ - *средняя генеральной совокупности;*

$\Delta_{\bar{x}}$ - *предельная ошибка выборки для средней.*

Предельная ошибка выборки:

$$\Delta_{\bar{x}} = t \cdot \sqrt{\frac{S^2}{n}\left(1 - \frac{n}{N}\right)}$$, *где*

t - *квантиль нормального закона распределения (при* α*=0,05;* t*=1,96);*

N - *объем генеральной совокупности;*

n - *объем выборки;*

S^2 - *исправленная выборочная дисперсия.*

Определение 3. ***Квантилем*** *или* ***нормированным отклонением*** *называется отношение предельной ошибки к средней ошибке.*

$$t = \frac{\Delta_{\bar{x}}}{M_{\bar{x}_0}}$$, *где* $M_{\bar{x}_0} = \frac{\sigma(x)}{\sqrt{n}}$

Квантиль вычисляется по соответствующему уровню значимости α *(при* $n \geq 30$, t *- квантиль нормального закона распределения, при* $n < 30$, t *- квантиль распределения Стьюдента).*

Существуют таблицы значений для ***t*** *в зависимости от уровня значимости* **α**.

Важной является задача определения объема выборочной совокупности ***n*** *при заданном уровне значимости. В случае бесповторного отбора необходимый объем выборки определяется по формуле:*

$$n = \frac{t^2 \cdot S^2 \cdot N}{t^2 \cdot S^2 + \Delta_{\bar{x}}^2 \cdot N}$$

Пример: *По условию задачи 2. При уровне значимости* **α=0,05** *определить:*

1) *несмещенные оценки математического ожидания, дисперсии и среднего квадратического отклонения;*

2) *доверительный интервал для математического ожидания с доверительной вероятностью* **(1-α)**;

3) *объем выборки, при котором с доверительной вероятностью* **(1-α)** *предельная ошибка выборки уменьшится в 2 раза, при сохранении уровня остальных характеристик.*

Учитывая, что проводилась 10% случайная бесповторная выборка.

Решение.

1) Несмещенной оценкой M(x) является выборочная средняя $\bar{x}$ $\bar{x} = 8{,}613$

Несмещенной оценкой D(x) является исправленная выборочная дисперсия S^2

$$S^2 = \sigma^2(x) \cdot \frac{n}{n-1} = \frac{5{,}981 \cdot 60}{59} = 6{,}082$$

Несмещенной оценкой σ(x) является стандартное отклонение S:

$$S = \sqrt{S^2} = \sqrt{6{,}082} = 2{,}466$$

2) Средняя численность работников на 100 га с/х угодий = 8,61. Найдем доверительный интервал для средней: $\bar{x} - \Delta_{\bar{x}} < \overline{x_0} < \bar{x} + \Delta_{\bar{x}}$

$\Delta_{\bar{x}} = t \cdot \sqrt{\frac{S^2}{n}\left(1 - \frac{n}{N}\right)}$ ***- предельная ошибка выборки для средней.***

При уровне значимости α=0,05 квантиль нормального распределения t=1,96.

Учитывая, что проводилась 10% выборка, N=10 ·60=600=>

$$\Delta_{\bar{x}} = 1{,}96 \cdot \sqrt{\frac{6{,}082}{60}\left(1 - \frac{60}{600}\right)} = 0{,}592$$

Значит, с доверительной вероятностью 1-α=0,95, можно утверждать, что средняя численность работников на 100 га с/х угодий во всей совокупности хозяйств находится в границах

$\bar{x} \pm \Delta_{\bar{x}} = 8{,}61 \pm 0{,}592$ ***, т.е. от 8,021 до 9,205.***

3) Необходимый объем выборки, для того, чтобы предельная ошибка не превышала $0{,}5 \cdot \Delta_{\bar{x}}$ ***при заданном уровне значимости*** **α=0,05** ***в случае случайного бесповторного отбора, определяется по формуле:***

$$n = \frac{t^2 \cdot S^2 \cdot N}{t^2 \cdot S^2 + \Delta_{\bar{x}}^2 \cdot N}$$

$$\Delta_{\bar{x}}^2 = \left(0{,}5 \cdot \Delta_{\bar{x}}\right)^2 = \left(0{,}5 \cdot 0{,}592\right)^2 = \left(0{,}296\right)^2$$

$$n = \frac{\left(1{,}96\right)^2 \cdot 6{,}082 \cdot 600}{\left(1{,}96\right)^2 \cdot 6{,}082 + \left(0{,}296\right)^2 \cdot 600} = \frac{14018{,}766}{23{,}365 + 52{,}576} = 185$$

Значит, для уменьшения предельной ошибки в два раза объем совокупности необходимо увеличить в 3 раза.

Математическая статистика

Глава 2. Элементы теории корреляции.

Известно, что процессы, протекающие в растениях и живых организмах, обусловлены влиянием большого числа взаимосвязанных факторов, среди которых имеются главные, определяющие основные свойства и характеристики процесса или явления, и второстепенные.

Как найти в виде формулы зависимость между двумя случайными величинами, полученными в результате наблюдений, если каждому значению одной величины соответствует несколько значений другой?

Как найти параметры этих формул при условии, чтобы они отражали сущность изучаемого процесса и «сглаживали» влияние случайных, не характерных для данного процесса факторов? Насколько сильно влияет изменение одной величины на изменение другой? Ответы на эти вопросы составляют содержание настоящей главы.

§1. Понятие о корреляции.

В сельскохозяйственных науках, в отличие от точных наук, полные (точные) функциональные связи встречаются редко, так как возможность искусственной изоляции влияния других факторов на изучаемые признаки в большинстве случаев неосуществима.

Например, связь урожайность - удобрения, имеется, но есть еще ряд факторов, влияющих на урожайность (севообороты, семена, предшественники, агротехника - субъективные факторы; метеорологические факторы- объективные).

Поэтому связь урожайность - удобрения неполная функциональная связь. Эту связь называют корреляционной (англ. correlation – соотношение, соответствие).

Метод корреляции применяется для того, чтобы при сложном взаимодействии посторонних влияний выяснить, какова была бы зависимость между результатом и фактором, если бы посторонние причины (факторы) не изменялись и своим изменением не искажали основную зависимость.

Первая задача корреляции: *выявление на основе наблюдений над большим количеством фактов того, как изменяется в среднем результативный признак в связи с изменением данного фактора (парная корреляция) или группы факторов (множественная корреляция). Эта задача решается нахождением уравнения связи.*

Вторая задача корреляции: *определение степени влияния искажающих факторов. Эта задача решается при помощи различных показателей тесноты связи: коэффициента корреляции, корреляционного отношения.*

Определение 1. ***Процесс нахождения связи между признаками называется выравниванием.***

Выравнивание ведет к нахождению переменной средней $\overline{y}_x$ ***, исчисленной в предположении функциональной зависимости*** ***y*** ***от*** ***x*****, т.е.** $\overline{y}_x = f(x)$ ***, и называется уравнением регрессии.***

При изучении влияния одних признаков на другие выделяются два признака - факториальный и результативный. Выделение этих признаков осуществляется путем логического анализа.

Например, в связи урожайность - осадки, урожайность - результативный признак, а осадки - факториальный.

Графическое изображение связи изучаемых явлений позволяет не только установить наличие или отсутствие связи между ними, но и изучить характер этой связи (форму связи и тесноту связи).

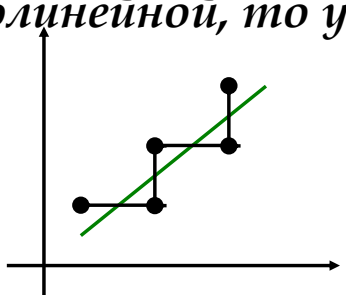
Если имеются числовые характеристики факториальных и результативных признаков одного и того же явления, то каждую пару чисел можно изобразить графически, откладывая по оси абсцисс - факториальный признак, по оси ординат - результативный признак.

§2. Графическое изображение связи.

Ломаная, соединяющая эти точки, называется ***ломаной регрессии.***

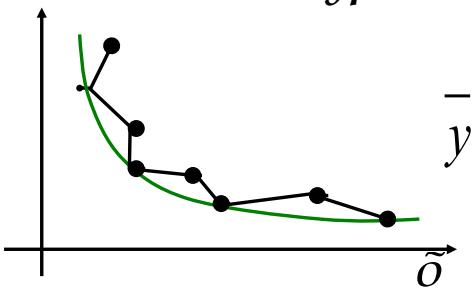
По форме этой ломаной приближенно определяют вид зависимости.

1) Если из графика видно, что связь близка к прямолинейной, то уравнение регрессии пишется в виде:



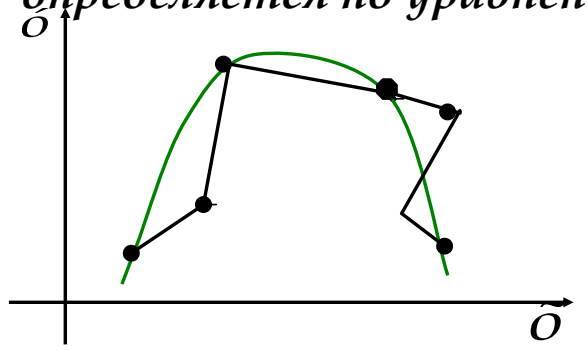
$$\overline{y} = ax + b$$

2) Если экспериментальные данные располагаются так, что через них можно провести гиперболу, то можно ожидать уравнение в виде:



$$\overline{y} = \frac{k}{x}; \quad \overline{y} = \frac{a}{x+b}, \quad \overline{y} = \frac{a}{x+b} + c$$

3) ***Если кривая имеет max или min, то зависимость определяется по уравнению:***



$$\overline{y} = ax^2 + bx + c$$

Для выявления функциональных зависимостей и определения неизвестных коэффициентов этой зависимости можно воспользоваться ***методом наименьших квадратов*****.**

$$\begin{cases} a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i\, y_i \\ a\sum_{i=1}^{n} x_i + b\cdot n = \sum_{i=1}^{n} y_i \end{cases} \Rightarrow \quad y = ax + b$$

Математическая статистика

Глава 2. Элементы теории корреляции.

§2. Графическое изображение связи.

$$\begin{cases} a\sum_{i=1}^{n} x_i^4 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i^2 \cdot y \\ a\sum_{i=1}^{n} x_i^3 + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i \cdot y_i \\ a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i + c \cdot n = \sum_{i=1}^{n} y_i \end{cases} \Rightarrow \quad y = ax^2 + bx + c$$

§3. Коэффициент корреляции.

После того, как уравнение регрессии найдено, находят так называемый коэффициент корреляции. Он используется для оценки тесноты связи между величинами при прямолинейной зависимости. Обозначается буквой r и определяется по формуле:

$$r = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\cdot\left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}\cdot\sqrt{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}}$$

, где

§3. Коэффициент корреляции.

$\bar{x}$ - среднее значение факториального (причинного)

признака $\bar{x} = \dfrac{\sum x_i}{n}$

$\bar{y}$ - среднее значение результативного

признака $\bar{y} = \dfrac{\sum y_i}{n}$

Промежуточные вычисления удобно располагать в виде таблицы:

№ наблюдения	x_i	y_i	$x_i - \bar{x}$	$\left(x_i - \bar{x}\right)^2$	$y_i - \bar{y}$	$\left(y_i - \bar{y}\right)^2$	$\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)$
$\sum$	…	…	…	…	…	…	…

Величина коэффициента корреляции находится в пределах $-1 \le r \le 1$ *:*

1) *Чем ближе* $|r|$ *к 1, тем теснее связь между факториальным и результативным признаками.*
2) *при* $|r|=1$ *получается полная функциональная связь.*
3) *если* $|r| \to 0$ *, то связь между признаками слабая.*

4) *при* $|r|=0$ *связи между признаками нет (линейная зависимость отсутствует).*
5) *при* $r>0$ *зависимость между признаками прямая (возрастающая).*
6) *при* $r<0$ *зависимость обратная (убывающая).*

Если зависимость между признаками прямая, то можно пользоваться уравнением прямой регрессии:

$$y - \bar{y} = b_{y/x}\left(x - \bar{x}\right)$$ *, где*

$b_{y/x}$ *- коэффициент регрессии, который определяется по формуле:*

$$b_{y/x} = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\cdot\left(y_i - \bar{y}\right)}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$$

Задача: *Для 10 петушков леггорнов 15 дневного возраста были получены следующие данные о весе их тела (x) в граммах и весе гребня (y) (в мг):*

x_i	83	72	69	90	90	95	95	91	75	70
y_i	56	42	18	84	56	107	90	68	31	48

Требуется:

1) *найти коэффициент корреляции и сделать вывод о тесноте и направлении линейной корреляционной связи между признаками;*
2) *составить уравнение прямой регрессии;*
3) *нанести на чертеж исходные данные и построить прямую регрессии.*

Решение:

Составим вспомогательную таблицу

№	x_i	y_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})(y_i-\bar{y})$
1	83	56	0	0	-4	16	0
2	72	42	-11	121	-18	324	198
3	69	18	-14	186	-42	1764	588
4	90	84	7	49	24	576	168
5	90	56	7	49	-4	16	-28
6	95	107	12	144	47	2209	564
7	95	90	12	144	30	900	360
8	91	68	8	64	8	64	64
9	75	31	-8	64	-29	841	232
10	70	48	-13	169	12	144	156
$\sum$	830	600	0	990	0	6854	2302

Вычисляем средние:

$$\bar{x}=\frac{\sum x_i}{n}=\frac{830}{10}=83 \qquad \bar{y}=\frac{\sum y_i}{n}=\frac{600}{10}=60$$

1) найдем коэффициент корреляции:

$$r=\frac{\sum(x_i-\bar{x})\cdot(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2}\cdot\sqrt{\sum(y_i-\bar{y})^2}} \quad r=\frac{2302}{\sqrt{990\cdot 6854}}=0,88$$

Вывод: ***между весом тела x и весом гребня y у 15-дневных петушков существует тесная положительная линейная корреляционная связь.***

2) найдем коэффициент регрессии:

$$b_{y/x} = \frac{\sum\left(x_i - \bar{x}\right)\cdot\left(y_i - \bar{y}\right)}{\sum\left(x_i - \bar{x}\right)^2} \qquad b_{y/x} = \frac{2302}{990} \approx 2,32$$

Подставим в уравнение прямой регрессии:

$$y - \bar{y} = b_{y/x}\left(x - \bar{x}\right)$$

$$y - 60 = 2,32\left(x - 83\right)$$

$$\underline{y = 2,32x - 132,56}$$

§4. Понятие о множественной корреляции

Если рассматривать зависимость результативного признака от двух или нескольких факториальных признаков, то придется изучать уравнения множественной корреляции (множественная связь).

В простейшем уравнении множественной связи предполагается, что зависимость между признаками линейная. Уравнение связи имеет вид: $z = a + bx + cy$.

При этом решаются задачи корреляции:

1. по данным наблюдений находится уравнение связи, т.е. определяются коэффициенты регрессии *a*, *b* и *c*;

2. оценивается теснота связи между *z*, *x* и *y*;

3. оценивается теснота связи между *z* и *x* (при постоянном *y*), между *z* и *y* (при постоянном *x*).

Первую задачу, т.е. определение параметров уравнения связи, решают методом наименьших квадратов с помощью системы нормальных уравнений:

$$\begin{cases} \sum y^2 \cdot c + \sum yx \cdot b + \sum y \cdot a = \sum yz \\ \sum xy \cdot c + \sum x^2 \cdot b + \sum x \cdot a = \sum xz \\ \sum y \cdot c + \sum x \cdot b + n \cdot a = \sum z \end{cases}$$

где, n – число одновременных наблюдений по трем признакам;

$\sum x, \sum y, \sum z$ - суммы соответствующих значений по этим признакам.

Пример: Рассмотрим зависимость урожайности ячменя (z) на одинаковых участках от количества внесенных минеральных удобрений (x) и количества выпавших в период цветения осадков (y). Данные наблюдений занесем в специальную таблицу:

z	x	y	zx	zy	xy	x^2	y^2	z_{xy}
0,8	20,3	10,0	16,24	8,00	203,00	412,09	100,00	0,84
0,9	21,7	10,1	19,53	9,09	219,17	470,89	102,01	0,96
1,0	20,5	10,6	20,50	10,60	217,30	420,25	112,36	0,97
1,1	21,8	10,2	23,98	11,22	222,36	475,24	104,04	1,00
1,2	23,1	10,9	27,72	13,08	251,79	533,61	118,81	1,21
1,3	23,6	11,2	30,68	14,56	264,32	556,96	125,44	1,31
6,3	131,0	63,0	138,65	66,55	1377,94	2869,04	662,66	6,29

Пользуясь данными таблицы, составим систему нормальных уравнений:

$$\begin{cases} 662{,}66c + 1377{,}94b + 63a = 66{,}55 \\ 1377{,}94c + 2869{,}04b + 131a = 138{,}65 \\ 63c + 131b + 6a = 6{,}3 \end{cases}$$

Решая эту систему получим:

$$c = 0{,}193; \quad b = 0{,}0711; \quad a = -2{,}5288,$$

и соответственно уравнение связи:

$$\overline{z_{xy}} = -2{,}5288 + 0{,}0711x + 0{,}0130y$$

$$\begin{array}{llll} x_1 = 20{,}3 & y_1 = 10{,}0 & \overline{z_{x_1y_1}} = 0{,}84 \\ x_2 = 21{,}7 & y_2 = 10{,}1 & \overline{z_{x_2y_2}} = 0{,}96 \\ x_3 = 20{,}5 & y_3 = 10{,}6 & \overline{z_{x_3y_3}} = 0{,}97 \\ x_4 = 21{,}8 & y_4 = 10{,}2 & \overline{z_{x_4y_4}} = 1{,}00 \\ x_5 = 23{,}1 & y_5 = 10{,}9 & \overline{z_{x_5y_5}} = 1{,}21 \\ x_6 = 23{,}6 & y_6 = 11{,}2 & \overline{z_{x_6y_6}} = 1{,}31 \end{array}$$

Как видно из расчетов, некоторые значения $\overline{z_{xy}}$ совпадают с опытными, а отдельные их значения мало отличаются друг от друга.

Но уравнение связи можно искать в виде:

$$z - \bar{z} = A(x - \bar{x}) + B(y - \bar{y}),$$

где, $$A = \frac{r_{xz} - r_{yz} \cdot r_{xy}}{1 - r_{xy}^2} \cdot \frac{\sigma_z}{\sigma_x}; \qquad B = \frac{r_{yz} - r_{xz} \cdot r_{xy}}{1 - r_{xy}^2} \cdot \frac{\sigma_z}{\sigma_y}.$$

Здесь r_{xz}, r_{yz}, r_{xy} - коэффициенты корреляции

между признаками x и y, x и z, y и z, σ_x, σ_y, σ_z - средние квадратические отклонения

$$r_{xy}=\frac{\overline{xy}-\bar{x}\cdot\bar{y}}{\sigma_x\cdot\sigma_y};\quad r_{xz}=\frac{\overline{xz}-\bar{x}\cdot\bar{z}}{\sigma_x\cdot\sigma_z};\quad r_{yz}=\frac{\overline{yz}-\bar{y}\cdot\bar{z}}{\sigma_y\cdot\sigma_z};$$

$$\sigma_x=\sqrt{\overline{x^2}-\bar{x}^2};\quad \sigma_y=\sqrt{\overline{y^2}-\bar{y}^2};\quad \sigma_z=\sqrt{\overline{z^2}-\bar{z}^2}.$$

Теснота связи признака z с признаками x, y оценивается выборочным совокупным коэффициентом корреляции:

$$R=\sqrt{\frac{r_{xz}^2-2r_{xy}\cdot r_{xz}\cdot r_{yz}+r_{yz}^2}{1-r_{xy}^2}},\qquad 0\le R\le 1,$$

при $R = 0$ линейная связь между x, y, z отсутствует, а при $R = 1$ между ними имеется точная математическая линейная связь $z=a+bx+cy$.

§5. Простейшие случаи криволинейной корреляции

O_1 Если график регрессии $\bar{y}_x=f(x)$ или $\bar{x}_y=\varphi(y)$ изображается кривой линией, то корреляцию называют криволинейной.

Пример: функция регрессии y на x могут иметь

вид:

1. $\bar{y}_x = ax^2 + bx + c$ - параболическая корреляция II порядка;

2. $\bar{y}_x = ax^3 + bx^2 + cx + d$ - параболическая корреляция III порядка.

Для определения вида функции регрессии строят точки (x_i и $\bar{y}_{x_i}$) и по их расположению делают заключение о примерном виде функции регрессии. При окончательном решении принимают во внимание особенности, вытекающие из сущности решаемой задачи.

Теория криволинейной корреляции решает те же задачи, что и теория линейной корреляции:

1. установление формы связи;

2. установление тесноты корреляционной связи.

Первая задача: рассмотрим параболическую корреляцию II порядка, предположив, что данные *n* наблюдений (выборки) позволяют считать, что имеет место именно такая

корреляция.

В этом случае выборочное уравнение регрессии имеет вид:

$\bar{y}_x = Ax^2 + Bx + C$,

где, A, B и C – неизвестные коэффициенты.

В этом случае, тоже пользуясь методом наименьших квадратов, можно получить систему нормальных уравнений для нахождения коэффициентов A, B и C.

$$\begin{cases} \sum n_x \cdot x^4 \cdot A + \sum n_x \cdot x^3 \cdot B + \sum n_x \cdot x^2 \cdot C = \sum n_x \cdot \bar{y}_x \cdot x^2; \\ \sum n_x \cdot x^3 \cdot A + \sum n_x \cdot x^2 \cdot B + \sum n_x \cdot x \cdot C = \sum n_x \cdot \bar{y}_x \cdot x; \\ \sum n_x \cdot x^2 \cdot A + \sum n_x \cdot x \cdot B + n \cdot C = \sum n_x \cdot \bar{y}_x. \end{cases}$$

Пример: найти выборочное уравнение регрессии $\bar{y}_x$ вида $\bar{y}_x = Ax^2 + Bx + C$ по данным корреляционной таблицы:

x / y	1	1,1	1,2	n_y
6	8	2	-	10
7	-	30	-	30

7,5	-	1	9	10
n_x	8	33	9	50

$$\bar{y}_{x_1} = \frac{6 \cdot 8}{8} = 6; \quad \bar{y}_{x_2} = \frac{6 \cdot 2 + 7 \cdot 30 + 7{,}5 \cdot 1}{33}; \quad \bar{y}_{x_3} = \frac{7{,}5 \cdot 9}{9} = 7{,}5.$$

Для получения системы нормальных уравнений составим вспомогательную таблицу:

x	n_x	$\bar{y}_x$	$n_x x$	$n_x x^2$	$n_x x^3$	$n_x x^4$	$n_x \bar{y}_x$	$n_x \bar{y}_x x$	$n_x \bar{y}_x x^2$
1	8	6	8	8	8	8	48	48	48
1,1	33	6,95	36,3	39,93	43,93	48,32	222,09	244,30	268,73
1,2	9	7,5	10,8	12,96	15,55	18,66	67,50	81,00	97,20
$\sum$	50	20,45	55,1	60,89	67,48	74,98	337,59	373,30	713,93

Используя суммы последней строки таблицы, составим систему нормальных уравнений:

$$\begin{cases} 74{,}98A + 67{,}48B + 60{,}89C = 413{,}93 \\ 67{,}48A + 60{,}89B + 55{,}1C = 373{,}30 \\ 60{,}89A + 55{,}1B + 50C = 337{,}59 \end{cases}$$

Решим систему и найдем:

$A = 1{,}94$; $B = 2{,}98$; $C = 1{,}10$.

Искомое уравнение регрессии:

$$\bar{y}_x = 1{,}94x^2 + 2{,}98x + 1{,}10.$$

Можно убедиться, что условные средние, вычисленные по этому уравнению, незначительно отличаются от условных средних корреляционной таблицы:

$x_1 = 1;$ $\bar{y}_{x_1} = 6;$

$x_2 = 1;$ $\bar{y}_{x_2} = 1{,}94 + 2{,}98 + 1{,}10 = 6{,}02$

таким образом, найденное уравнение хорошо согласуется с данными выборки.

Вторая задача: теснота связи.

Для линейной корреляции теснота связи характеризовалась выборочным коэффициентом корреляции, для криволинейной корреляции теснота связи характеризуется корреляционным отношением: η (тэта).

$\eta_{y/x}$ - выборочное корреляционное отношение y на x;

$\eta_{x/y}$ - выборочное корреляционное отношение x на y;

O_1 Выборочным корреляционным отношением y на x называется отношение межгруппового среднего квадратического отклонения к общему квадратическому отклонению признака y.

$$\eta_{y/x} = \frac{\sigma_{\text{межгр.}}}{\sigma_{\text{общ.}}} = \frac{\sigma_{\bar{y}x}}{\bar{\sigma}_y}; \qquad \sigma_{\bar{y}x} = \sqrt{D_{\text{межгр.}}} = \sqrt{\frac{\sum_{i=1}^{n} n_x (\bar{y}_x - \bar{y})^2}{n}};$$

$$\bar{\sigma}_y = \sqrt{D_{\text{общ.}}} = \sqrt{\frac{\sum_{i=1}^{n} n_y (y - \bar{y})^2}{n}},$$

n – объем выборки;

n_x – частота значения признака x;

n_y – частота значения признака y;

$\bar{y}$ – общая средняя признака y;

$\bar{y}_x$ – условная средняя признака y.

Аналогично находится и $\eta_{x/y}$.

$$\eta_{x/y} = \frac{\sigma_{\bar{x}y}}{\bar{\sigma}_x}.$$

Свойства корреляционного отношения:

1. корреляционное отношение удовлетворяет неравенству $0 \le \eta \le 1$;

2. если $\eta = 0$, то признак y с признаком x корреляционно не связаны;

3. если $\eta = 1$, то y связан с x функциональной зависимостью;

4. $\eta \geq |r_B|$ – выборочное корреляционное отношение не меньше абсолютной величины выборочного коэффициента корреляции;

5. если $\eta = |r_B|$, то имеет место точная линейная корреляционная зависимость.

ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

§1. Общая постановка задачи

O_1 Статистическая гипотеза – это предположение о виде и параметрах распределения случайной величины x.

Проверка гипотезы производится на основе выборочного наблюдения и в зависимости от того, в какой мере выборочные данные согласованы с выдвинутым предположением:

гипотеза либо подтверждается, либо отвергается.

O_2 Гипотеза, которая подлежит проверке, называется основной или нулевой и обозначается H_0.

Для любой основной гипотезы может существовать одна или несколько альтернативных теорий (конкурирующих). Основная гипотеза часто состоит в предположении об отсутствии существенных расхождений (о нулевом расхождении) между *ожидаемым* и *фактическим* значениями параметра.

При проверке гипотез возможны ошибки двух видов:

1. ошибка I рода – гипотеза H_0 отвергнута, когда она истинна;

2. ошибка II рода – гипотеза H_0 принята, когда в действительности она не верна.

Таким образом, здесь возможны 4 события, вероятности которых зависят от *исходных*

данных, от используемого *критерия* и заданного уровня *значимости*.

Результаты проверки	Оценка принятого решения, если	
	верна H_0	верна альтернативная гипотеза
Гипотеза H_0 отклоняется	ошибка I рода	правильное решение
Гипотеза H_0 не отклоняется	правильное решение	ошибка II рода

O_3 Статистический критерий – это правило, в соответствии с которым принимается или отклоняется гипотеза. Построение критерия сводится к выбору подходящей функции $T(x)$, которая называется статистикой критерия. Значения наиболее употребляемых критериев табулированы и приведены в специальных таблицах $T_{кр.}$, $T_{табл.}$, T_x; их сравнивают с $T_{факт.}$ или $T_{набл.}$. Уровень значимости определяет критическую область статистического критерия.

1. если $T_{факт.} > T_{кр.}$, то расхождения значимы и H_0

отвергается.

2. Если $T_{факт.} \leq T_{кр.}$, то гипотеза H_0 принимается.

O_4 Уровень значимости равен вероятности совершить ошибку I рода. Надежность критерия повышается при уменьшении x, но параллельно происходит снижение мощности критерия (вероятности недопущения ошибки II рода). Единственным способом уменьшить вероятность совершить ошибки I и II рода является увеличение объема выборки. Принятие ложной гипотезы более зло, чем отклонение верной. В экономических и сельскохозяйственных исследованиях чаще всего используют уровни значимости 0,05 и 0,01.

§2. Классификация статистических гипотез

Разнообразие статических критериев можно классифицировать по трем признакам:

1. характер (цель) решаемой задачи;

2. расположение критической области;

3. использование информации о виде распределения.

1. В зависимости от характера или цели решаемой задачи используются критерии проверки и критерии согласия.

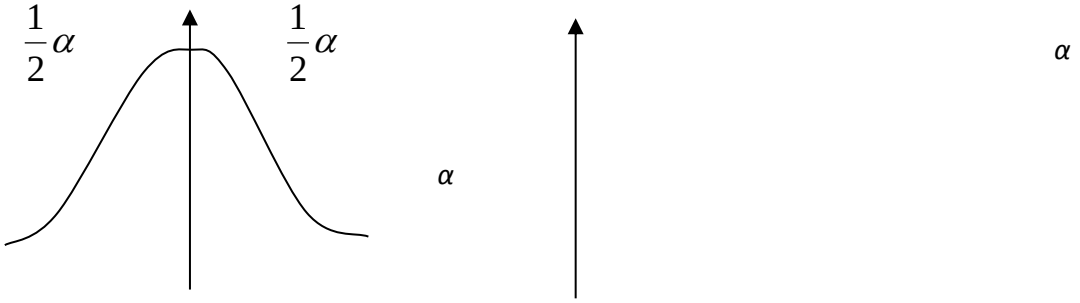
Критерии проверки – это статистические критерии, служащие для проверки гипотез относительно отдельных параметров генеральной совокупности.

Для построения критериев проверки используют распределение Стьюдента (t – распределение), нормальное распределение, F – статистика Фишера-Снедекора.

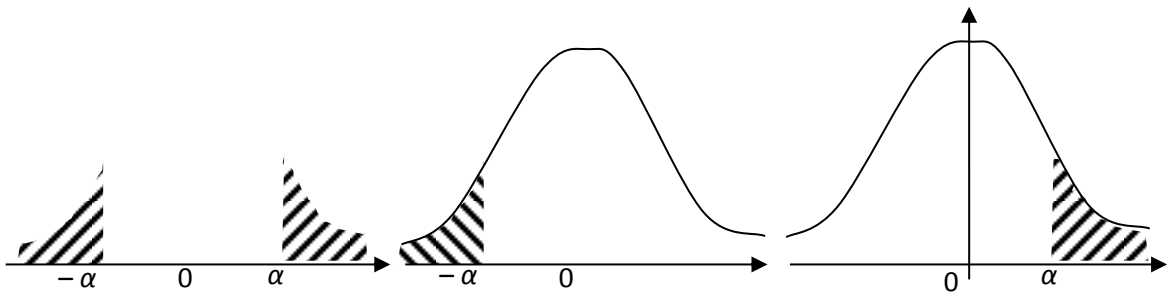
Критерии согласия применяются при проверке гипотез о соответствии эмпирических наблюдений теоретическим. Это критерий χ^2 (хи) Пирсона, критерий Колмогорова.

2. По расположению критической области критерии проверки делятся на односторонние и двухсторонние критерии.

Пример: имеются 2 выборочные ($\tilde{\chi}_1$ и $\tilde{\chi}_2$). Нулевая гипотеза H_0 состоит в предположении, что



разность между этими выборочными $(\alpha = \tilde{\chi}_1 - \tilde{\chi}_2)$ несущественна $(\alpha = 0)$. Тогда альтернативные гипотезы будут: $H\alpha_1 - (\alpha \neq 0)$, $H\alpha_2 - (\alpha > 0)$, $H\alpha_3 - (\alpha < 0)$.



Двусторонняя критическая область (H_0)

Левосторонняя критическая область $(H\alpha_2)$

Правосторонняя критическая область $(H\alpha_3)$

3. В зависимости от того, привлекается или нет информация о виде случайных величин (в форме оценок параметров для построения статистических критериев), среди них выделяют параметрические и непараметрические.

Параметрические (классические) основаны на том предположении, что распределение исходных данных подчинено одному из известных законов, чаще всего нормальному. К ним относятся: F – распределение Фишера-Снедекора, t – распределение Стьюдента, χ^2 Пирсона и другие.

Непараметрические критерии строятся непосредственно по данным выборочного наблюдения. Они используются, если распределение признака генеральной совокупности неизвестно, а ее параметры не оценены. Число таких критериев велико. К ним относится критерий Колмогорова.

§3. Проверка гипотез относительно доли признака

1. Сравнение доли признака с нормативом.

Пусть: a – нормативная величина;

W_r – доля признака в генеральной совокупности;

H_0 – нулевая гипотеза, состоит в утверждении H_0:

$W_r - a = 0$.

Используется двусторонний критерий.

Доля признака в выборке (W_B) при любом значении n распределена по биноминальному закону:

для повторного отбора:

$$P_{(x=m)} - C_n^m \cdot p^m \cdot q^{n-m};$$

для бесповторного отбора по гипергеометрическому закону:

$$P_{(x=m)} = \frac{C_M^m \cdot C_{N-M}^{n-m}}{C_M^n} \cdot p^m \cdot q^{n-m},$$

где, N – объем генеральной совокупности;

n – объем выборки;

M – число значений случайной величины с данным признаком в генеральной совокупности;

m – число значений с данным признаком в выборке.

При больших значениях n вычисления могут быть упрощены аппроксимацией (приближением) нормальным распределением

или если W_r достаточно мала, то используют распределение Пуассона.

Нормальная аппроксимация.

Значением критерия t может быть получено по формуле:

$$t_{\text{набл.}} = \frac{|W - a|}{\sqrt{\frac{a(1-a)}{n}}}.$$

Табличное значение t_α при уровне значимости α находят из равенства $2\Phi(t) = 1 - \alpha$ $(P = 1 - \alpha)$ (см. таблицу значений $\Phi(t)$ – интегральная функция Лапласа).

Если $t_{\text{набл}} > t_\alpha$, то H_0 отвергается. Если же $t_{\text{набл.}} \le t_\alpha$, то есть основание считать, что данные выборки согласуются с нулевой гипотезой.

Нулевую гипотезу H_0 можно проверить иначе: найти интеграл $B_{\alpha,u} = a \pm t_\alpha \cdot \sqrt{\frac{a(1-a)}{n}}$. Если при этом выполняется неравенство $B_\alpha \le W \le B_u$, то H_0 не отвергается. В противном случае H_0 отклоняется.

Пример 1. В хозяйстве получено 700 голов приплода КРС, из них 364 бычка. Требуется

проверить гипотезу о том, что отклонение доли бычков в общем поголовье приплода от норматива a=0,5 случайно (уровне значимости α=0,05).

Решение: Применим двусторонний критерий, то есть будем считать, что $W-a=0 \quad (H_0: W-a=0)$.

Оценим долю бычков в поголовье по данным наблюдений

$$W=\frac{364}{700}=052.$$

Вычислим наблюдаемое значение критерия по формуле:

$$t_{\text{набл}}=\frac{|W-a|}{\sqrt{\frac{a(1-a)}{n}}}$$

$$t_{\text{набл}}=\frac{|0{,}52-0{,}5|}{\sqrt{\frac{0{,}5(1-0{,}5)}{700}}}=\frac{0{,}02\cdot 10\cdot\sqrt{7}}{0{,}5}\approx 1{,}06$$

Табличное значение t_α:

$$2\Phi(t_\alpha)=1-\alpha;$$
$$2\Phi(t_\alpha)=0{,}95$$
$$\Phi(t_\alpha)=0{,}475$$
$$t_\alpha=1{,}96$$

Так как $t_{\text{набл}}<t_\alpha \quad (1{,}06<1{,}96)$, то H_0 не отвергается.

Аналогично расчет можно провести и для интервальной оценки:

$$B_{\alpha,u} = a \pm t_{\alpha} \cdot \sqrt{\frac{a(1-a)}{n}}$$

$$B_{\alpha,u} = 0{,}5 \pm 1{,}96 \cdot \sqrt{\frac{0{,}5(1-0{,}5)}{700}}$$

$$0{,}463 \le W_B \le 0{,}537$$

$$0{,}463 \le 0{,}5 \le 0{,}537$$

2. Сравнение долей признака в двух совокупностях.

Пусть есть 2 выборки из одной генеральной совокупности с выборочными долями соответственно W_1 и W_2. Выдвинем H_0, что W_1 и W_2 расходятся несущественно, то есть $\alpha = W_1 - W_2$ находится в пределах возможных ошибок выборочного наблюдения, или (иначе) W_1 и W_2 являются хорошими оценками доли в генеральной совокупности (W_r), а это послужит основанием считать нулевую гипотезу верной.

Так как W_r неизвестно, то приходится использовать ее оценку, получаемую из

объединенных данных:

$W = \frac{m_1 + m_2}{n_1 + n_2}$, где

n_1 - объем первой выборки;

n_2 - объем второй выборки;

m_1 - число элементов 1 выборки, обладающих данными признаками;

m_2 - число элементов 2 выборки, обладающих тем же признаком.

Оценка *дисперсии доли* в W_r вычисляется по формуле:

$$S^2 = \frac{W(1-W)}{n}.$$

Оценка выборочной дисперсии разности долей:

$$S_\alpha^2 = S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right) = \frac{S^2(n_1 + n_2)}{n_1 \cdot n_2}.$$

Для проверки нулевой гипотезы $H_0 : \alpha = 0$ возможно использование t – критерия с двусторонней критической областью. И так:

$$t_{\text{набл}} = \frac{|\alpha|}{\sqrt{W(1-W)}} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}. \qquad (1)$$

Можно иначе:

$$\begin{cases} D(x \pm y) = D(x + y) \\ \sigma^2_{x-y} = \sigma^2_x + \sigma^2_y \end{cases}$$

$$\sigma^2_{W_1 - W_2} = \sigma^2_{W_1} + \sigma^2_{W_2}$$

И тогда: $S^2_\alpha = \frac{W_1(1-W_1)}{n_1} + \frac{W_2(1-W_2)}{n_2}$.

В этом случае критерий

$$t_{\text{набл.}} = \frac{|\alpha|}{\sqrt{\frac{W_1(1-W_1)}{n_1} + \frac{W_2(1-W_2)}{n_2}}}. \qquad (2)$$

Пример 2: В базисном году в области существовало 30 самостоятельных фермерских хозяйств при общем количестве сельскохозяйственных предприятий 250. В отчетном году число фермерских хозяйств увеличилось до 60 единиц, а сельскохозяйственных предприятий до 280.

Требуется установить, насколько существенными являются изменения доли фермерских хозяйств.

Решение:

Пусть W_1 и W_2 - доли фермерских хозяйств в

базисном и отчетном годах, $H_0 : W_1 - W_2 = \alpha = 0$.

Альтернативная гипотеза $H_a : W_2 > W_1$.

Пусть $\alpha = 0{,}05$.

$$\left.\begin{array}{l} m_1 = 30 \\ m_2 = 60 \\ n_1 = 250 \\ n_2 = 280 \end{array}\right| \begin{array}{l} W_1 = \dfrac{m_1}{n_1} = \dfrac{30}{250} = 0{,}12 \\ W_1 = \dfrac{m_2}{n_2} = \dfrac{60}{280} = 0{,}214 \end{array}$$

$$W = \frac{m_1 + m_2}{n_1 + n_2} = \frac{30 + 60}{250 + 280} = \frac{90}{530} = 0{,}17$$

$$|\alpha| = |W_1 - W_2| = |0{,}12 - 0{,}214| = |-0{,}094| = 0{,}094$$

$$t_{\text{набл.}} = \frac{0{,}094}{0{,}17(1-0{,}17)} : \sqrt{\frac{250 + 280}{250 \cdot 280}} = 2{,}88$$

(1)

$$t_{\text{набл.}} = \frac{0{,}094}{\sqrt{\frac{0{,}12(1-0{,}12)}{250} + \frac{0{,}214(1-0{,}214)}{280}}} = 2{,}94 \qquad (2)$$

При уровне значимости $\alpha = 0{,}05$ по таблице $t = 1{,}96$ (см. предыдущий пример).

Так как $t_{\text{набл.}} > t_\alpha$ $(2{,}88; 2{,}94 > 1{,}96)$, следовательно H_0 отвергается, предпочтение отдается $H_a : W_2 > W_1$.

§4. Проверка гипотез относительно распределений

Если известен закон распределения изучаемого признака, тогда легко проводить статистический

анализ. Проверить гипотезу о соответствии эмпирического распределения любому теоретическому закону можно с помощью критериев согласия.

Критерий χ^2 Пирсона:

Наблюдаемое значение критерия вычисляется по формуле:

$$\chi^2_{\text{набл.}} = \sum_{i=1}^{n} \frac{\left(m_i - m_i^*\right)^2}{m_i^*},$$

где m_i – эмпирические частоты распределения;

m_i^* – теоретические частоты.

Табличное значение критерия находят по заданному уровню значимости α и числу степеней свободы $\gamma = k - b - 1$, где:

k – число групп (интервалов) эмпирического ряда распределения;

b – число параметров теоретического распределения.

При проверки нормальности (подчиненности нормальному закону) эмпирического ряда

распределения $b = 2$, так как в нормальном распределении два параметра $\tilde{x} = \alpha; \sigma$. Тогда $\lambda = k - 3$.

Для кривой нормального распределения частоты вычисляются по формуле:

$$m_i^* = \frac{n \cdot h}{\sigma} \cdot \varphi(t_i),$$

где n – объем выборки;

h – величина интервала (в интервальном ряду) или положительная разность между соседними значениями признака (в дискретном ряду);

σ – среднее квадратическое отклонение признака x;

$\varphi(t_i)$ – значение функции Лапласа $\varphi(t)$.

$t = \frac{x - \tilde{x}_B}{\sigma}$ (см. прилож. $\varphi(t)$).

Если фактические частоты незначительно отличаются от теоретических или совпадают с ними $\left(\chi^2_{\text{набл.}} \le \chi^2_{\text{теор(крит.)}}\right)$, то гипотеза о соответствии эмпирического ряда теоретическому не отвергается. При существующем расхождении

фактических и теоретических частот $\chi^2_{\text{набл.}} > \chi^2_{\text{теорет.}}$ и H_0 (гипотезу о соответствии эмпирического теоретическому) отвергают.

Пример 3: Дано распределение 60 хозяйств области по обеспеченности зерноуборочными комбайнами Дон-1500.

x_i (число комбайнов)	1	2	3	4	5	6
m_i (количество хозяйств)	5	20	16	9	6	4

$n = 60$.

Проверить при уровне значимости $\alpha = 0{,}05$ гипотезу о соответствии распределения хозяйств в области по обеспеченности комбайнами нормальному закону.

По данным выборки уже найдено: $h = 1;\quad \tilde{x}_B = 3;\quad \sigma = 1{,}35$.

Дальнейшие расчеты теоретических частот проведем в таблице:

x_i	m_i	$\varphi(t_i)$	$t_i = \frac{x_i - 3}{1{,}35}$	$m_i^* = \frac{60 \cdot 1}{1{,}35} \cdot \varphi(t_i)$
1	5	0,1334	-1,48	6
2	20	0,3034	-0,74	13

3	16	0,3989	0,00	18
4	9	0,3034	0,74	13
5	6	0,1334	1,48	6
6	4	0,0508	2,03	2
∑	60			58

где: x_i – фактические значения признака;

t_i – нормированное отклонение;

$\varphi(t_i)$ – функция Лапласа;

m_i^* –теоретические частоты.

При правильно выполненных расчетах сумма теоретических частот $(\sum m_i^*)$ равна или незначительно отличается от объема выборки $(\sum m_i = n)$, что имеет место в этом примере.

Теперь вычислим наблюдаемое значение χ^2

m_i	m_i^*	$m_i - m_i^*$	$(m_i - m_i^*)^2$	$\frac{(m_i - m_i^*)^2}{m_i^2}$
5	6	-1	1	$\frac{1}{6} = 0,17$
20	13	7	49	$\frac{49}{13} = 3,77$
16	18	-2	4	0,22
9	13	-4	16	1,23
6	6	0	0	0,00
4	2	2	4	2,00

60	58			7,39

$\chi^2_{\text{набл.}} = 7{,}39$.

Для нахождения табличного значения χ^2: $\alpha = 0{,}05;\ \gamma = 6 - 3 = 3$ – по таблице χ^2 (см. приложение). $\chi^2_{\text{крит.}} = 7{,}81$.

Так как $\chi^2_{\text{набл.}}(7{,}39) < \chi^2_{\text{крит.}}(7{,}81)$, то нулевая гипотеза H_0 не отвергается.

Замечание: при вычислении $\chi^2_{\text{набл.}}$ возможны ситуации, когда приходится делить на теоретические частоты $m_i^* \approx 0$ или $m_i^* = 0$, если округляют до целых. Чтобы этого избежать, следует объединить крайние группы так, что наименьшая теоретическая частота получится больше или равна единице (к такому процессу тоже подходят аккуратно, так как это ведет к уменьшению числа степеней свободы).

§5. Сравнение дисперсий

Дисперсия (σ^2) является одним из основных

показателей вариации (рассеивание) признака. К сравнению дисперсий прибегают тогда, когда требуется сопоставить совокупности по степени их однородности.

Проверка статистических гипотез относительно дисперсий имеет важное значение и для принятия экономических решений.

Пример: Сравнительная экономическая оценка двух сортов сельскохозяйственной культуры. Основными биологическими признаками сортов являются:

1. средняя многолетняя урожайность;
2. количество зерен;
3. вариация урожайности в связи с действием различных факторов.

При равенстве первых показателей, последний становится главным в принятии решения об использовании сортов.

Для случайных различных вариаций урожайности возможны такие варианты

решения:

1. предпочтение отдается сорту, имеющему меньшую колеблемость урожайности по годам, так как при этом требуется меньше соответствующей техники и рабочей силы для уборки урожая без потерь;

2. предпочтение отдается сорту, имеющему большую вариацию урожайности, поскольку выяснено, что эта вариация связана с действием факторов, уровни которых планируется регулировать;

3. выбор – за оптимальным сочетанием посевов двух сортов, учитывающим особенности каждого из них и поставленную экономическую цель.

Эти методы измерения вариаций оценивания действия факторов опираются на расчет и сравнение дисперсий. Так, сопоставление вариации признака в двух совокупностях достигается проверкой гипотезы о равенстве дисперсий. Для этого применяется *F*-критерий

Фишера-Снедекора.

Наблюдаемые значения критерия вычисляются по формуле:

$F_{\text{набл.}} = \frac{\sigma_1^2}{\sigma_2^2}$, где $\sigma_1^2 \geq \sigma_2^2$.

$F_{\text{крит.}}$ находят из таблицы F-распределения (см. приложение) по заданному уровню значимости α и числу степеней свободы γ_1 и γ_2 соответственно для большой и малой дисперсий:

$\gamma_1 = n_1 - 1;$ $\gamma_2 = n_2 - 1.$

Если $F_{\text{набл.}} \geq F_{\text{кр.}}$, то гипотеза о равенстве дисперсий отвергается, а если наоборот $F_{\text{набл.}} < F_{\text{кр.}}$, то гипотеза не отвергается.

Использование F-критерия предполагается, что обе совокупности имеют нормальное распределение.

Пример 4: В результате испытания двух сортов гороха на трех сортоучастках за пять лет получены следующие сведения:

A	$Б$
$\widetilde{x}_B = 19{,}8$	$\widetilde{x}_B = 20$, ц/га
$\sigma_2^2 = 9{,}98$	$\sigma_1^2 = 25{,}12$

Требуется оценить, существенно ли различие в вариации урожайности сортов.

Так как урожайности близки друг к другу, то для ответа на этот вопрос достаточно проверить гипотезу о равенстве дисперсий.

По исходным данным

$$F_{\text{набл.}} = \frac{\sigma_1^2}{\sigma_2^2} = \frac{25{,}12}{9{,}98} = 2{,}51.$$

Критическое значение критерия при $\alpha = 0{,}05$ и $n_1 = n_2 = 5 \cdot 3 = 15$; $\gamma_1 = 15 - 1 = 14$; $\gamma_2 = 15 - 1 = 14$ равно $F_{\text{кр.}} = 2{,}48$.

Так как $F_{\text{набл.}}(2{,}51) > F_{\text{кр.}}(2{,}48)$, то гипотеза о равенстве дисперсий отвергается с риском ошибочности в 5 случаях ил 100 ($\alpha = 0{,}05$).

Замечание: $F_{\text{кр.}} = 2{,}48$ получено из таблицы

приложения F - распределения интерполяцией (чтением между строк) данных.

γ	12	14
14	2,53	2,35

Вывод: есть основание полагать, что урожайность сорта *Б* менее стабильна в связи с различными погодными и почвенными условиями.