Лекция.2. ВЕРОЯТНОСТНЫЕ МОДЕЛИ ОДНОМЕРНОЙ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ

Вопросы лекции:

1. **Случайные события, величины, вероятности**
2. **Функция распределения вероятности и функция плотности вероятности – как количественные модели генеральной совокупности**
3. **Числовые характеристики генеральных совокупностей**

До проведения исследований основные свойства генеральной совокупности являются неизвестными, а сами исследования предназначены для установления этих свойств. Поэтому адекватными являются вероятностные представления, основу которых составляет понятие случайных событий, величин, вероятностей и их распределений.

Вопрос 1. Случайные события, величины, вероятности

Случайное событие – такое событие, которое в результате некоторого эксперимента (наблюдения) может появиться или нет

Например, выпадение «орла» при подбрасывании монеты, возникновение пожара при страховании имущества.

Событие появляется или нет при выполнении определенного комплекса условий.

- Если выполнение некоторого комплекса условий приводит к однозначному результату (появление или не появление некоторого события), то такое событие называется детерминированным (вполне определённым);

- если при неизменных условиях, которые не полностью определяют поведение некоторого события, регистрируются факты его появления в цепи экспериментов 1, 2,3, …,n, то можно получить некоторое количество $\nu A(n)$ этих фактов. Рассмотрим отношение: $P{\sim}A(n) = \nu A(n)/n$ (отношение появления исследуемого события к общему числу экспериментов)

Очевидно, что выполняется неравенство:

$$0 \leq P{\sim}\, A(n) \qquad (2.1)$$

Если $n \to \infty$, тогда $P{\sim}A(n)$ стремится к некоторому пределу PA, который называется вероятностью, P

Вероятность - является неизвестной и часто даже неизвестно, существует ли этот предел. Поэтому существование вероятности некоторого события в виде определённого числа удовлетворяющего неравенству:

$$PA = P, \; 0 \le P \le 1 \qquad (2.2)$$

является одной из гипотез, используемых при статистическом анализе.

(*Например, из 1000 застрахованных объектов строительства за прошлый год аварии произошли на 50 объектах. Тогда вероятность аварийности, рассчитываемая для принятия мер по страхованию объектов, составит: P=50/1000)*

Случайная величина – переменная, принимающая значение из некоторого диапазона числовой оси [a, b].

Обычно рассматривают событие попадания значений случайной величины в некоторый интервал числовой оси.

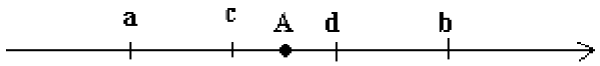

Рис.2.1. Случайная величина

Тогда событие А означает, что $\varepsilon \in [c, d]$, где ε - случайная величина. С этими событиями связана вероятность $PA = P\ \{\varepsilon \in [c, d]\}$. (то есть вероятность того, что случайная величина попадет в промежуток от с до d. Самый простой пример – рулетка, попадание шарика в определенное поле)

Основной проблемой является поиск такого интервала минимальной длины, когда выполняется условие:

$$\begin{aligned} &|d - c| \to \min \\ &PA = P\ \{\varepsilon \in [c, d]\} = P \end{aligned} \qquad (2.2)$$

(то есть мы только с определенной вероятностью, а не наверняка можем предсказать, каким будет значение случайной величины, например, какова будет прибыль в будущем году)

При предсказании значений случайной величины ошибки предсказания характеризуются двумя параметрами:

-Шириной задаваемого интервала возможных значений, которые характеризуют абсолютную погрешность предсказания (*Наша задача – как можно точнее предсказать, например, прибыль будет от 0 до 100000руб – очень широкий интервал, а от 95000 до 99000 – уже приемлемый*);

-вероятностью попадания значения случайной величины в этот интервал, которая характеризует надёжность прогноза (*например, прибыль равную 95000 мы получим с вероятностью 90%. Точно предсказать никогда не возможно из-за множества действующих в экономике факторов(перечислите факторы, влияющие на прибыль)*).

Понятие случайной величины формулируется в теории вероятностей и является вероятностной моделью генеральной совокупности (*например, если мы проведем опрос и выясним, что третья часть вашей студенческой учебной группы помнит теорию вероятности, то совсем не факт, что из всей генеральной совокупности студентов МГСУ третья часть помнит теорию вероятности. Относительно всех студентов доля помнящих* *теорию вероятности будет случайной величиной, а относительно выборки – вашей группы – эта доля будет детерминированной, то есть определенной*).

Вопрос 2. Функция распределения вероятности и функция плотности вероятности – как количественные модели генеральной совокупности

Вероятности событий заключающихся в попадании значений случайной величины на некоторые интервалы числовой оси, чаще всего вычисляются с помощью функции плотности вероятности (*ФПВ*) или функции распределения вероятности (*ФРВ*).

Закон распределения дискретной случайной величины представляет собой перечень всех её возможных значений и соответствующих вероятностей. Сумма всех вероятностей $\Sigma p_i = 1$. (*например, при подбрасывании монеты возможны два варианта событий – «орел» и «решка». Они оба равновероятны: вероятность выпадения «орла» равна вероятности выпадения «решки» и равна 0,5. Тогда сумма всех вероятностей: 0,5+0,5=1)*

Функция распределения случайной величины - это вероятность того, что случайная величина (назовём её ξ) примет значение меньшее, чем конкретное числовое значение x:

$$F(X) = P(\xi < X). \qquad (2.3)$$

Для дискретной случайной величины функция распределения вычисляется для каждого значения как сумма вероятностей, соответствующих всем предшествующим значениям случайной величины.

<u>Примеры некоторых непрерывных распределений</u>

Нормальное распределение

Нормальное распределение имеет плотность вероятности $1/[\sigma\sqrt{2\pi}]\cdot e^{-(x-a)2/2\sigma2}$, где a - математическое ожидание, σ - среднее квадратическое отклонение.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}} \qquad (2.4)$$

(Значения плотности нормального распределения для конкретного числового значения x можно вычислить в Excel с помощью формулы =НОРМРАСП(x;a;σ;0). Если a = 0, σ = 1, то такое нормальное распределение называется стандартным. Значения плотности стандартного нормального распределения можно посмотреть в таблице или вычислить в Excel с помощью формулы =НОРМРАСП(x;0; 1;0)) Иногда нормальную кривую называют кривой Гаусса.

При изменении параметра *a* форма нормальной кривой не изменяется. В этом случае, если математическое ожидание (параметр *a)* уменьшилось или увеличилось, график нормальной кривой сдвигается влево или вправо .

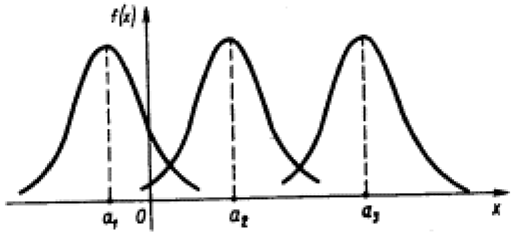

Рис.2.2. График нормальной кривой при изменении параметра а

При изменении параметра σ изменяется форма нормальной кривой. Если этот параметр увеличивается, то максимальное значение функции *f(x)* убывает, и наоборот. Так как площадь, ограниченная кривой распределения и осью *Ox,* должна быть постоянной и равной 1, то с увеличением параметра кривая приближается к оси *Ox* и растягивается вдоль нее, а с уменьшением σ кривая стягивается к прямой $x=a$.

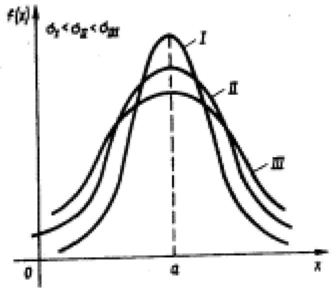

Рис.2.3. График нормальной кривой при изменении параметра σ

Использование формул f(x) и F(x) для практических расчетов затруднительно. Но решение задач по этим формулам можно упростить, если от нормального распределения с произвольными параметрами *a* и σ перейти к нормальному распределению с параметрами a=0, σ = 1.

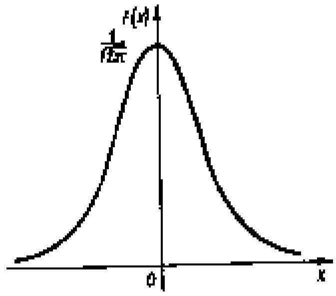

Рис.2.4. График нормальной кривой с параметрами а=0, $\sigma = 1$

Для вычисления вероятности попадания случайной величины в интервал (α, β) воспользуемся функцией Лапласа

Тогда

$$P(\alpha < X < \beta) = P(\frac{\alpha - a}{\sigma} < \frac{X - a}{\sigma} < \frac{\beta - a}{\sigma}) = P(u_1, u_2) = \Phi(u_2) - \Phi(u_1). \qquad (2.5)$$

Значения функции Ф(u) необходимо взять из таблицы "Таблица значений функции Ф(х)"из любого учебника статистики или теории вероятностей .

***Пример**. Случайная величина X распределена по нормальному закону. Математическое ожидание и среднее квадратическое отклонение этой величины соответственно равны 30 и 10. Найти вероятность того, что X примет значение, принадлежащее интервалу (10, 50).*

Решение:

По условию:α =10, β=50, a=30, σ =10, следовательно,

$$P(10 < X < 50) = \Phi\left(\frac{50-30}{10}\right) - \Phi\left(\frac{10-30}{10}\right) = 2\Phi(2) \qquad (2.6)$$

По таблице находим Ф (2) = 0,4772. Отсюда, искомая вероятность:

P(10 < X < 50) =2·0,4772=0,9544.

Правило трех сигм:

вероятность того, что отклонение по абсолютной величине будет меньше утроенного среднего квадратического отклонения, равна 0,9973.

Другими словами, вероятность того, что абсолютная величина отклонения превысит утроенное среднее квадратическое отклонение, очень мала, а именно равна 0,0027=1-0,9973. Это означает, что лишь в 0,27% случаев так может произойти. Такие события, исходя из принципа невозможности маловероятных событий, можно считать практически невозможными. В этом и состоит сущность правила трех сигм.

На практике правило трех сигм применяют так: если распределение изучаемой случайной величины неизвестно, но условие, указанное в приведенном правиле, выполняется, то есть основание предполагать, что изучаемая величина распределена нормально; в противном случае она не распределена нормально.

Пример практического применения закона нормального распределения случайной величины для решения конкретной экономической задачи.

Требуется произвести изучение интенсивности движения автомобилей на участке платной автомобильной дороги с целью обоснования эффективности инвестиционного проекта строительства автодороги на платной основе. Расчет произвести в системе электронных таблиц Excell с использованием инструмента «Генерация случайных чисел». Данный инструмент дает возможность исследовать генеральную совокупность при имеющихся данных о выборочной совокупности. Пример применения инструмента «Генерация случайных чисел» для моделирования распределения объемов грузоперевозок по Ростовской области в 2001 году приведен на рисунке(здесь, согласно выборке, определена средняя *интенсивность движения 10019авт/сут, стандартное отклонение по выборке 1850) .*

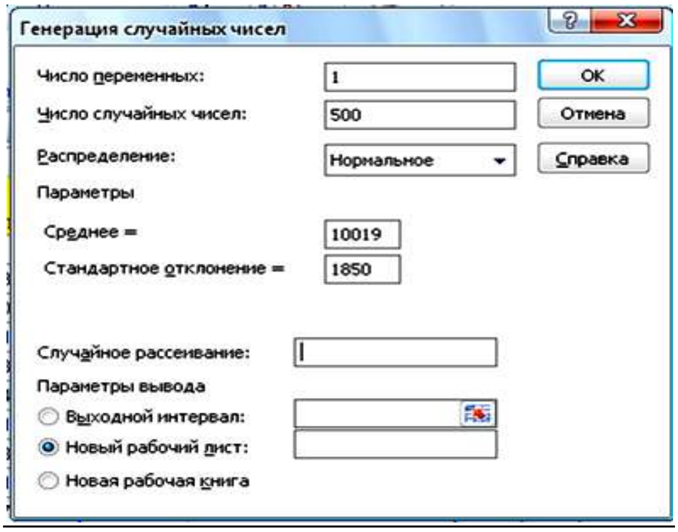

Рис2.5. Заполнение полей диалогового окна «Генерация случайных чисел»

Для осуществления имитации считаем целесообразным использовать нормальное распределение, так как практика показала, что именно оно встречается в подавляющем большинстве случаев. Количество имитаций может быть сколь угодно большим и определяется требуемой точностью анализа. В данном случае расчет ограничен 500 имитациями.

Таблица 2.1 – Результаты построения массива прогнозной интенсивности движения на платном участке 1070-1091 км автомагистрали М-4 «Дон» в Ростовской области в 2011 году

№ п. п.	Интенсивность движения, авт./сут.
1	8337,159
2	8728,122
3	10335,6
4	12393,65
5	9468,079
6	11594,67
7	12273,01
8	13374,98
9	10555,88
10	10589,76
…	И т. д. 500 имитаций

На основе полученных в результате имитации данных, используя инструмент «Описательная статистика» MS Excel проведен экономико-статистический анализ.

Таблица 2.2 – Оценка статистических критериев расчетного массива прогнозной интенсивности движения на платном участке 1070-1091 км автомагистрали М-4 «Дон» в Ростовской области в 2011 году

Наименование критерия	Значение
Среднее	10870,53
Стандартная ошибка	83,56632

Медиана	*10712,16*
Мода	*10605,04*
Стандартное отклонение	*1868,6*
Дисперсия выборки	*3491665*
Эксцесс	*-0,11867*
Асимметричность	*0,131174*
Интервал	*11033,24*
Минимум	*6151,899*
Максимум	*17185,14*
Сумма	*5435266*
Счет	*500*

Осуществим оценку значимости коэффициента асимметрии для распределения интенсивности движения. Наиболее простым способом получения такой оценки является определение стандартной (средней квадратической) ошибки асимметрии, рассчитываемой по формуле:

$$\sigma_{as} = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}} \qquad (2.7)$$

где n *- число значений случайной величины (в данном случае -* 500).

Если отношение коэффициента асимметрии s *к величине ошибки* σ_{as} *меньше трех (т.е.:* s / σ_{as} *< 3), то асимметрия считается несущественной, а ее наличие объясняется воздействием случайных факторов*

Аналогичным способом можно осуществить проверку значимости величины эксцесса – е . Формула для расчета стандартной ошибки эксцесса имеет следующий вид:

$$\sigma_{ex} = \sqrt{\frac{24n(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}} \qquad (2.8)$$

где:

n *- число значений случайной величины.*

Если отношение e / σ_{ex} *< 3, эксцесс считается незначительным и его величиной можно пренебречь.*

В результате экономико-статистического анализа прогноза интенсивности движения на платном участке 1070-1091 км автомагистрали М-4 «Дон» в Ростовской области в 2011 году получено среднее прогнозное значение интенсивности 10870,53±83,56 авт./сут. Стандартное отклонение 1868,6. Значение коэффициента вариации 0,17, что говорит о достаточно большом риске неточности прогноза интенсивности. Минимальное значение интенсивности на участке 6151,9 авт./сут., максимальное – 17185,14 авт./сут. В соответствии с правилом *«трех сигм», вероятность интенсивности движения в интервале [10870,53-1868,6; 10870,53+1868,6] равна 0,68.*

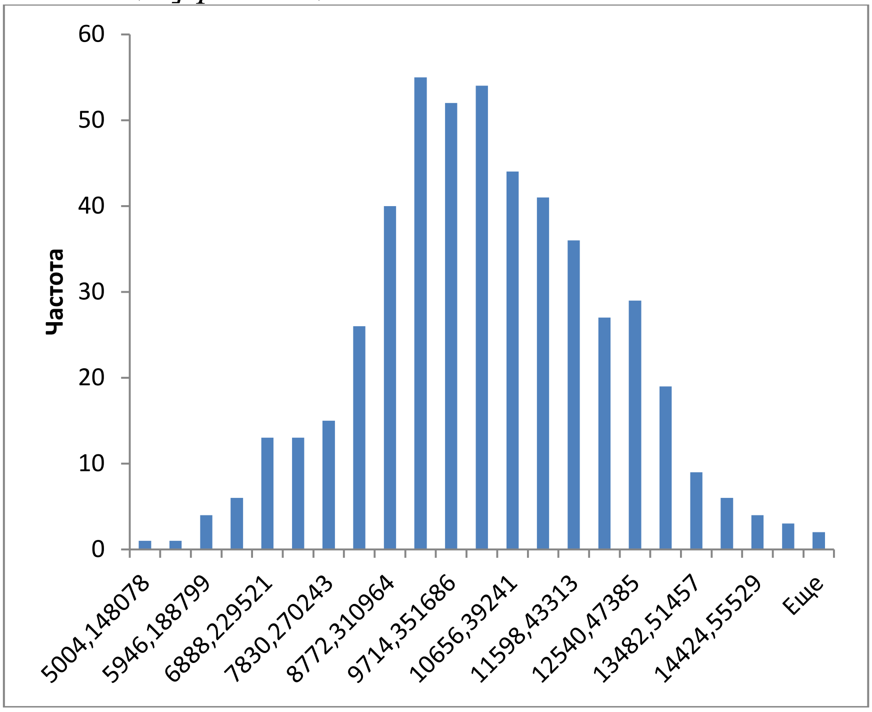

Рис. 2.6. Распределение значений интенсивности движения (гистограмма частот)

Как видим, фактическая гистограмма частот имеет вид кривой нормального распределения с небольшой асимметрией и скосом.

Равномерное распределение

Плотность вероятности равномерного распределения сохраняет на интервале (a, b) постоянное значение, вне этого интервала плотность вероятности равна нулю. Исходя из основного свойства плотности вероятности,

$$f(x) = 1/(b-a) \tag{2.9}$$

на интервале (a;b)

Пример применения показательного распределения к решению прикладных задач.

Автобусы некоторого маршрута идут строго по расписанию. Интервал движения 5 минут. Найти вероятность того, что пассажир, подошедший к остановке. Будет ожидать очередной автобус менее 3 минут.

Решение:

СВ- время ожидания автобуса имеет равномерное распределение. Тогда искомая вероятность будет равна: $P(0,3) = \dfrac{3-0}{5-0} = 0,6.$

Показательное распределение.

Непрерывная случайная величина X, функция плотности которой задается выражением

$$f(x) = \begin{cases} \mu e^{-\mu x}, x > 0; \\ 0, x \le 0, \end{cases} \tag{2.10}$$

называется случайной величиной, имеющей показательное, или экспоненциальное, распределение.

Величина срока службы различных устройств и времени безотказной работы отдельных элементов этих устройств при выполнении определенных условий обычно подчиняется показательному распределению. Другими словами, величина промежутка времени между появлениями двух последовательных редких событий подчиняется зачастую показательному распределению.

Как видно из формулы , показательное распределение определяется только одним параметром μ.

График функции показательного распределения имеет вид:

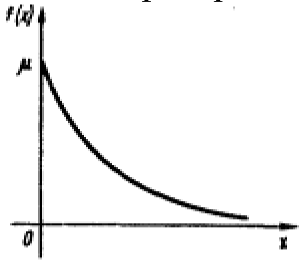

Рис. 2.7. График функции показательного распределения

Вопрос 3. Числовые характеристики генеральных совокупностей

Основными числовыми характеристиками случайной величины являются следующие:

Мода – точка на числовой оси, в которой функция плотности вероятности принимает локальное или глобальное экстремальное значение типа “максимум”. (*это значение случайной величины, имеющее наибольшую вероятность. На многоугольнике распределения мода - это абсцисса самой высокой точки. Бывает, что распределение имеет не одну моду.)*

Распределение вероятности с одной модой называется унимодальным, в противном случае – полимодальным.

Положение моды (максимум) определяет участки повышенной вероятности попадания в них значений случайной величины

Медиана – точка, которая определяет два интервала (слева и справа), вероятности попадания в которые значения случайной величины одинаковы и равны 0,5.

Существует еще ряд критериев, описывающих вероятность наступления события и, главное, позволяющих определить величину его последствий. К таким критериям относятся:

- **Дисперсия** – сумма квадратов отклонений случайной величины от ее среднего значения, взвешенных на соответствующие вероятности.

$$\mathrm{Var}(R) = \sum_{i=1}^{n} p_i \cdot (X_i - M(R))^2 \qquad (2.11)$$

где p_i – вероятность появления значения X_i;

M(R) – среднее или ожидаемое значение (математическое ожидание) дискретной случайной величины *R* определяется как сумма произведений ее значений на их вероятности:

$$M(R)=\sum_{i=1}^{n}X_i\cdot p_i \qquad (2.12)$$

- **Математическое ожидание** – важнейшая характеристика случайной величины, т.к. служит центром распределения ее вероятностей. Смысл ее заключается в том, что она показывает наиболее правдоподобное значение фактора.

- На практике результаты анализа более наглядны, если показатель разброса случайной величины выражен в тех же единицах измерения, что и сама случайная величина. Для этих целей используют стандартное (среднее квадратическое) отклонение σ(R).

$$\sigma(R)=\sqrt{Var(R)} \qquad (2.13)$$

- Все вышеперечисленные показатели обладают одним общим недостатком – это абсолютные показатели, значения которых предопределяют абсолютные значения исходного фактора. Гораздо удобней поэтому использовать **коэффициент вариации** *(CV)*.

$$CV=\frac{\sigma(R)}{M(R)} \qquad (2.14)$$

Определение *CV* особенно наглядно для случаев, когда средние величины случайного события существенно различаются.

Пример применения вероятностных критериев к решению прикладных экономических задач.

Необходимо определить ключевые факторы риска жилищного строительства. Для этого предлагается применять анализ чувствительности по основным факторам риска, влияющим на финансовый результат проектов жилищного строительства в условиях экономического кризиса. Основными факторами, влияющими на динамику цен на рынке жилья, в настоящее время являются внешние по отношению к данному рынку факторы: общая экономическая ситуация в стране, ее особенности в отдельных регионах, привлекательность рынка недвижимости для капиталов, образующихся вследствие роста нефтяных доходов, по сравнению с другими рынками.

В качестве ключевых выбираются те факторы, изменения которых приводят к наибольшим отклонениям прибыли (что соответствует максимуму дисперсии).

Таблица 2.3 - Выбор ключевых факторов риска на основе анализа чувствительности

Факторы	*-20%*	*-15%*	*-10%*	*-5%*	*0*	*+5 %*	*+10%*	*+15%*	*+20%*	*Дисперсия прибыли*
Колебание спроса	*17749,79*	*18646,92*	*19589,40*	*20579,51*	*22239,78*	*23363,85*	*24544,73*	*25785,30*	*27088,57*	*30683926,61*
Колебание цены нефти	*32643,90*	*34141,73*	*35616,80*	*37070,71*	*38504,85*	*39920,46*	*41318,65*	*42700,39*	*44066,61*	*15275930,28*
Факторы	*-20%*	*-15%*	*-10%*	*-5%*	*0*	*+5 %*	*+10%*	*+15%*	*+20%*	*Дисперсия прибыли*
Колебания курса доллара	*28958,97*	*28959,30*	*28959,63*	*28959,96*	*28960,32*	*28960,65*	*28960,98*	*28961,31*	*28961,64*	*0,84*
Изменение ставки по кредитам	*24746,55*	*27750,86*	*30916,69*	*34243,03*	*37728,94*	*41373,51*	*45175,90*	*49135,29*	*53250,93*	*95490131,47*
Изменение доли ипотечных сделок	*27507,42*	*28464,23*	*29396,79*	*30307,02*	*31196,59*	*32066,97*	*32919,45*	*33755,20*	*34575,25*	*5844577,01*
Изменение объема инвестиций в основной капитал	*23458,47*	*24944,92*	*26431,38*	*27917,83*	*29404,29*	*30890,74*	*32377,20*	*33863,65*	*35350,11*	*16571610,16*
Динамика заработной платы	*29333,41*	*31148,44*	*32963,46*	*34778,49*	*36593,51*	*38408,54*	*40223,56*	*42038,59*	*43853,62*	*24707381,39*

По показателю дисперсии можно сделать вывод о наличии трех ключевых факторов риска: изменение ставки по кредитам, изменение объема инвестиций в основной капитал и изменение спроса на жилье.

С помощью анализа чувствительности можно также оценивать наиболее вероятную прибыль предприятия. В данном случае определены возможные значения прибыли от реализации 1 ^{м2} жилья с учетом вероятности появления этих значений, в результате чего получается массив значений прибыли.

Таблица 2.4 - Массив значений прибыли от реализации жилья

Вероятности	*0,027*	*0,2295*	*0,0135*	*0,039*	*0,104*	*0,117*	*0,188*	*0,235*	*0,047*
Прибыль	*23363,85*	*18646,92*	*13271,81*	*53250,93*	*59273,92*	*84283,28*	*33536,63*	*33120,42*	*31188,03*
Матем.ожи-дание	*630,82*	*4279,47*	*179,17*	*2076,79*	*6164,49*	*9861,14*	*6304,89*	*7783,30*	*1465,84*

В данном случае математическое ожидание выступает как взвешенная с учетом вероятности оценка возможной прибыли. Среднее ожидаемое значение прибыли равно сумме математических ожиданий (средневзвешенное значение с учетом вероятности)=38881,75 руб.

На основе данных массива прибыли можно рассчитать критерии устойчивости оценок финансового результата проектов жилищного строительства.

Таблица 2.5 - Определение критериев устойчивости оценок финансового результата проектов жилищного строительства

Ситуация	*1*			*2*			*3*		
Сценарии	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Средняя прибыль	*38881,75*								
Квадраты разностей	*240805359,2*	*347707702*	*176140946*	*2835661113*	*3513397457*	*7103670782*	*1124705651*	*1096962494*	*972693391,2*
Ст. отклон.	*20438,48*								
К-т вариации	*0,53*								
P(П<=0)	*0,03*								
P(П< = среднее)	*0,50*								
P(П>=среднее)	*0,50*								
P(П>макс)	*0,01*								
P(П> среднее +10%)	*0,42*								

Вероятности определялись с использованием функции НОРМРАСП системы электронных таблиц. Возвращает нормальную функцию распределения для указанного среднего и стандартного отклонения. Эта функция имеет очень широкий круг приложений в статистике, включая проверку гипотез.

Функции распределения (вероятности) определяют вероятность, что случайная величина будет принимать значение, меньшее или равное определенной величине.

На основании расчетов, приведенных в таблице, можно сделать вывод о том, что наиболее вероятная прибыль от вложения активов в строительство 1 м^2 жилья составит 38881,75 руб. Так как коэффициент вариации достаточно велик, можно сделать вывод о низкой устойчивости оценок эффективности данного вида активов. Причем несомненным фактором риска выступает рост ставок по кредитам и снижение спроса.

*Цена риска данного вида активов в соответствии с правилом «трех сигм» составит 3*20438,48=61315,43 руб. , что в достаточной степени превышает наиболее вероятную прибыль. Полученные значения коэффициента вариации и цены риска свидетельствует о низкой устойчивости оценок финансового результата проектов жилищного строительства. На основании значения коэффициента вариации можно сказать, что на 1 рубль среднего дохода приходится 53 копейки потерь.*

Контрольные вопросы по теме 2:

1. Приведите пример случайных величин и случайных событий.
2. Чем характеризуется разброс значений случайной величины?
3. Приведите примеры случайных величин, распределенных согласно различным законам.