Лекция №13-14.

Тема 8.10. Проверка статистических гипотез.

Основные распределения математической статистики: распределения Стьюдента, Пирсона, Фишера. Понятие статистической гипотезы. Проверка гипотезы о законе распределения. Доверительные интервалы для математического ожидания и дисперсии.

§ 5. Статистическая проверка статистических гипотез

Сопоставление высказанной гипотезы относительно генеральной совокупности с имеющимися выборочными данными, сопровождаемое количественной оценкой степени достоверности получаемого вывода и осуществляемое с помощью того или иного статистического критерия, называется проверкой статистических гипотез.

Определение. *Статистической гипотезой* называют гипотезу о виде неизвестного распределения генеральной совокупности или о параметрах известных распределений.

Определение. *Нулевой (основной)* называют выдвинутую гипотезу H_0. *Конкурирующей (альтернативной)* называют гипотезу H_1, которая противоречит нулевой.

Если выдвигаемая гипотеза сводится к утверждению о том, что значение некоторого неизвестного параметра генеральной совокупности в точности равно заданной величине, то эта гипотеза называется простой. В других случаях гипотеза называется сложной.

Выдвинутая нулевая гипотеза может быть правильной или неправильной, поэтому возникает необходимость статистической проверки справедливости основной гипотезы H_0 (табл.11).

Так как проверка статистических гипотез осуществляется на основании выборочных данных, то решение неизбежно сопровождается вероятностью ошибочного заключения как в ту, так и в другую сторону.

В какой-то небольшой доле случаев нулевая гипотеза может оказаться отвергнутой, в то время как она справедлива. Такую ошибку называют *ошибкой первого рода*, а её вероятность – *уровнем значимости* α. Другими словами это та вероятность, которой можно пренебречь.

Наоборот, в какой-то небольшой доле случаев нулевая гипотеза принимается, в то время как на самом деле она ошибочна. Такую ошибку называют *ошибкой второго рода*. Вероятность ошибки второго рода p. Вероятность $1-p$ называют мощностью критерия.

Таблица 11

Нулевая гипотеза H_0	Результаты решения относительно H_0	
	Отклонена	Принята
Верна	Ошибка 1-го рода, её вероятность $P(H_1/H_0) = \alpha$	Правильное решение, его вероятность $P(H_0/H_0) = 1-\alpha$

Неверна	Правильное решение, его вероятность $P(H_1/H_1) = 1 - p$	Ошибка 2-го рода, её вероятность $P(H_0/H_1) = p$

Для проверки нулевой гипотезы пользуются специально подобранной случайной величиной, распределение которой известно. В общем случае её обозначают K – критерий согласия, устанавливающий, когда полученное в действительности указанное отклонение следует принять несущественным, а когда существенным. Критерий K является функцией от результатов наблюдения.

Пример. Пусть H_0 заключается в том, что математическое ожидание генеральной совокупности $a = 3$. Тогда возможные варианты H_1: а) $a \neq 3$; б) $a > 3$; в) $a < 3$.

Основной прием проверки статистических гипотез заключается в том, что по имеющейся выборке вычисляется значение некоторой случайной величины, имеющей известный закон распределения.

Определение. *Статистическим критерием* называется случайная величина K с известным законом распределения, служащая для проверки нулевой гипотезы.

Определение. *Критической областью* называют область значений критерия, при которых нулевую гипотезу отвергают, *областью принятия гипотезы* – область значений критерия, при которых гипотезу принимают.

Поскольку критерий K – одномерная случайная величина, все ее возможные значения принадлежат некоторому интервалу. Поэтому критическая область и область принятия гипотезы также являются интервалами, следовательно, существуют точки, которые их разделяют.

Критическими точками (границами) $k_{кр}$ называют точки, отделяющие критическую область от области принятия гипотезы.

Различают одностороннюю (правостороннюю или левостороннюю) и двустороннюю критические области.

Правосторонней называют критическую область, определяемую неравенством $K > k_{кр}$, где $k_{кр}$ – положительное.

Левосторонней называют критическую область, определяемую неравенством $K < k_{кр}$, где $k_{кр}$ – отрицательное число.

Односторонней называют правостороннюю или левостороннюю критическую область.

Двусторонней называют критическую область, определяемую неравенствами $K < k_1$; $K > k_2$, где $k_2 > k_1$.

Для отыскания критической области, как было сказано выше, задаются уровнем значимости α. Критическую точку находят исходя из требования, чтобы при условии справедливости нулевой гипотезы выполнялись равенства:

а) для правосторонней критической области

$$P(K > k_{кр}) = \alpha \quad (k_{кр} > 0);$$

б) для левосторонней критической области

$$P(K < k_{кр}) = \alpha \quad (k_{кр} < 0);$$

в) для двусторонней симметричной области

$$P(K > k_{кр}) = (\alpha/2)\ (k_{кр} > 0),\ P(K < -k_{кр}) = (\alpha/2).$$

Итак, процесс проверки гипотезы состоит из следующих этапов:

1. Выбирается статистический критерий K.
2. Вычисляется его значение $K_{выб}$ по имеющейся выборке.
3. Поскольку закон распределения K известен, определяется (по известному уровню значимости) *критическое значение* $K_{кр}$, разделяющее критическую область и область принятия гипотезы (например, если $P(K > K_{кр}) = \alpha$, то справа от $K_{кр}$ располагается критическая область, а слева – область принятия гипотезы).
4. Если вычисленное значение $K_{выб}$ попадает в область принятия гипотезы, то нулевая гипотеза принимается, если в критическую область – нулевая гипотеза отвергается.

§ 6. Критерий Пирсона для проверки гипотезы о виде закона распределения случайной величины

С помощью критерия Пирсона можно проверять гипотезы о различных законах распределения. Для данного критерия в качестве величины, характеризующей степень различия между теоретическим и выборочным законами распределения, выбирается статистика

$$K = \chi^2 = \sum_{i=1}^{l} \frac{\left(n_i - n_i^*\right)^2}{n_i},$$

которая учитывает расхождения между теоретическими n_i и выборочными n_i^* частотами. Эта случайная величина называется χ^2 («хи квадрат») статистикой Пирсона. Смысл ее очевиден: суммируются квадраты отклонений эмпирических частот от теоретических. Целью исследования является сравнение эмпирических и теоретических частот, которые, конечно, отличаются друг от друга, и выяснить, являются ли эти различия несущественными, не опровергающими гипотезу о выбранном законе распределении исследуемой случайной величины, или они настолько велики, что противоречат этой гипотезе.

Можно доказать, что вне зависимости от реального закона распределения генеральной совокупности, закон распределения случайной величины при $n \to \infty$ стремится к закону распределения с числом степеней свободы $k = l - 1 - r$, где r – число параметров предполагаемого распределения, оцененных по данным выборки, l-число интервалов. Для выбранного критерия строится критическая область, определяемая условием

$$P\left(\chi_{\text{выб.}}{}^2 < \chi^2_{\text{крит}}(l;\alpha)\right) = \alpha,$$

где α – уровень значимости. Следовательно, критическая область задается неравенством $\chi_{\text{выб.}}{}^2 < \chi^2_{\text{крит}}(l;\alpha)$ а область принятия гипотезы $\chi_{\text{выб.}}{}^2 < \chi^2_{\text{крит}}(l;\alpha)$.

Таким образом, для проверки нулевой гипотезы H_0 *генеральная совокупность распределена по выбранному закону* – нужно вычислить по выборке наблюдаемое значение критерия:

$$\chi^2_{\text{выб.}} = \sum_{i=1}^{l} \frac{\left(n_i - n_i^*\right)^2}{n_i},$$

а по таблице критических точек распределения χ^2 найти критическую точку $\chi^2_{крит}(l;\alpha)$, используя известные значения α и $k = l - 1 - r$.

Если $\chi_{выб.}{}^2 < \chi^2_{крит}(k;\alpha)$ нулевую гипотезу принимают, при $\chi_{выб.}{}^2 > \chi^2_{крит}(k;\alpha)$ ее отвергают.

Замечание. Число параметров распределения r, оцененных по данным выборки для нормального и равномерного распределений равно двум, а для показательного – одному.

Из вышесказанного следует, что для поверки данной гипотезы H_0 необходимо найти теоретические и эмпирические частоты.

Пусть получена выборка достаточно большого объема n с большим количеством различных значений вариант. Для удобства её обработки разделим интервал от наименьшего до наибольшего из значений вариант на k равных частей по методике, предложенной в § 2. Будем считать, что значения вариант, попавших в каждый интервал, приближенно равны числу, задающему середину интервала. Подсчитав число вариант, попавших в каждый интервал, составим так называемую сгруппированную выборку (табл. 12).

Таблица 12

варианты $\bar{x}_i$	$\tilde{x}_1$	$\tilde{x}_1$	…		$\tilde{x}_k$
частоты n_i	n_1	n_2	…		n_k

Примечание. $\tilde{x}_i = \dfrac{x_i + x_{i+1}}{2}$ – значения середин интервалов, а n_i – число вариант, попавших в i-й интервал (эмпирические частоты).

По полученным данным можно вычислить выборочное среднее $\bar{x}_в$ и выборочное среднее квадратическое отклонение $\sigma_в$. Тип закона определим по построенной гистограмме или из общих соображений. Например, ошибки измерений в основном распределены по нормальному закону, а надежность безотказной работы прибора – по показательному. Для проверки предположения, что генеральная совокупность распределена по выбранному закону с параметрами $M[X] = \bar{x}_в$; $D[X] = \sigma_в^2$, необходимо вычислить теоретические частоты по формуле

$$n_i = p_i \cdot n,$$

где $p_i = p_i\left(x_i < X < x_{i+1}\right) = \int\limits_{x_i}^{x_{i+1}} f(x)dx$ – вероятность попадания в i-й интервал , n – объем выборки.

Здесь и далее $f(x)$ - функция плотности распределения вероятностей случайной величины *X*, выбранного закона распределения. Для простоты вычислений можно воспользоваться приближенной формулой вычисления определенного интеграла

$$p_i = p_i\left(x_i < X < x_{i+1}\right) = \int_{x_i}^{x_{i+1}} f(x)dx = f\left(\tilde{x}_i\right)\cdot\left(x_{i+1} - x_i\right) = f\left(\tilde{x}_i\right)\cdot h,$$

где $\tilde{x}_i = \dfrac{x_i + x_{i+1}}{2}$.

Рассмотрим практические приёмы нахождения теоретических вероятностей и теоретических частот для основных законов распределения.

1. *Нормальный закон распределения.*

Функция плотности распределения задается формулой

$$f(x) = \frac{1}{\sigma_в\sqrt{2\pi}}\cdot e^{\frac{-(x-\bar{x}_в)^2}{2\sigma_в{}^2}} = \frac{1}{\sigma_в}\cdot\varphi\left(\frac{x-\bar{x}_в}{\sigma_в}\right).$$

График функции $f(x)$ называют нормальной кривой или кривой Гаусса (рис. 3.3).

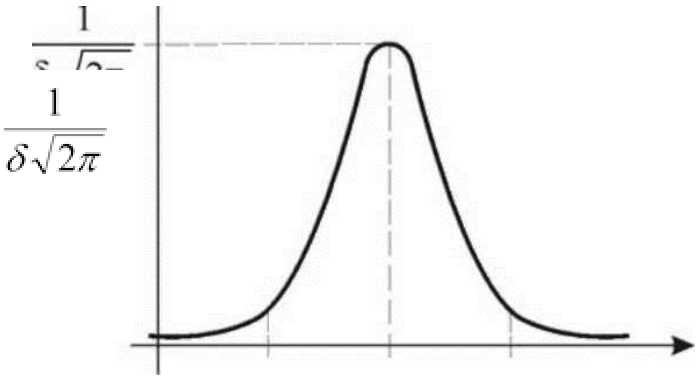

Рис. 3.3. Кривая Гаусса

Тогда вероятности p_i вычисляются по формуле

$$p_i = p_i\left(x_i < X < x_{i+1}\right) = \int_{x_i}^{x_{i+1}} f(x)dx = \Phi\left(\frac{x_{i+1} - \bar{x}_в}{\sigma_в}\right) - \Phi\left(\frac{x_i - \bar{x}_в}{\sigma_в}\right),$$

где $\Phi(x) = \dfrac{1}{2\pi}\int_0^x e^{-\frac{z^2}{2}}dz$ – функция Лапласа, её значения находят по прил. 2.

Чаще всего пользуются приближенной формулой

$$p_i = p_i\left(x_i < X < x_{i+1}\right) = \int_{x_i}^{x_{i+1}} f(x)dx = \frac{1}{\sigma_в}\cdot\varphi\left(\frac{\tilde{x}_i - \bar{x}_в}{\sigma_в}\right)\cdot\varDelta,$$

где $\varphi(x) = \dfrac{1}{\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2}}$ – табулированная функция .

Умножив полученные вероятности на объем выборки *n*, найдем теоретические частоты: $n_i = p_i \cdot n$.

2. *Равномерный закон распределения*.

При использовании критерия Пирсона для проверки гипотезы о равномерном распределении генеральной совокупности с предполагаемой плотностью вероятности

$$f(x)=\begin{cases}0, & x\notin[a,b];\\ \dfrac{1}{b-a}, & x\in[a,b]\end{cases}$$

необходимо, вычислив по имеющейся выборке значения $\bar{x}_в$ и $\sigma_в^2$, оценить параметры a и b.

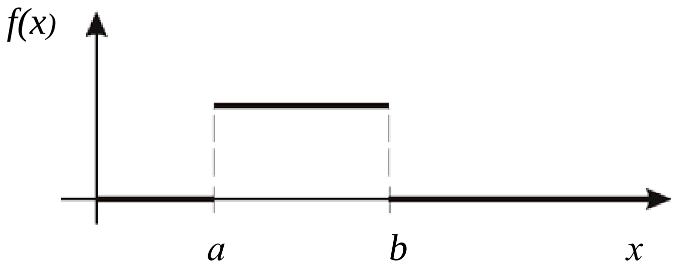

Рис.. График функции плотности равномерного распределения

Действительно, так как для равномерного распределения $M(X)=\dfrac{a+b}{2}$ и $\sigma(X)=\sqrt{D(X)}=\sqrt{\dfrac{(a-b)^2}{12}}=\dfrac{a-b}{2\sqrt{3}}$, то оценки параметров a и b – концов интервала, в котором наблюдались возможные значения случайной величины X, находим из системы (3.10), где через a^* и b^* обозначены оценки соответствующих параметров равномерного распределения.

$$\begin{cases}\dfrac{b^*+a^*}{2}=\bar{x}_в;\\ \dfrac{b^*-a^*}{2\sqrt{3}}=\sigma_в.\end{cases}$$

Решая систему уравнений, получим

$$a^*=\bar{x}_в-\sqrt{3}\sigma_в,\ b^*=\bar{x}_в+\sqrt{3}\sigma_в.$$

Тогда теоретическое распределение будет иметь вид

$$f(x)=\begin{cases}0, & x\notin[a^*,b^*];\\ \dfrac{1}{b^*-a^*}, & x\in[a^*,b^*].\end{cases}$$

Вероятности p_i вычисляются по формуле:

$$p_i=p_i(x_i<X<x_{i+1})=\int_{x_i}^{x_{i+1}}f(x)dx=\int_{x_i}^{x_{i+1}}\frac{1}{b^*-a^*}dx=\frac{x_{i+1}-x_i}{b^*-a^*}=\frac{\Delta}{b^*-a^*}.$$

Умножив полученные вероятности на объем выборки n, найдем теоретические частоты: $n_i=p_i\cdot n$.

3. *Показательный закон распределения*.

Функция плотности определяется формулой

$$f(x)=\begin{cases}0, x<0;\\ \lambda e^{-\lambda x}, x\geq 0.\end{cases}$$

Вычислив по имеющейся выборке значения $\bar{x}_в$ и $\sigma_в^2$, необходимо оценить параметр λ. Для показательного распределения $M(X)=\frac{1}{\lambda}$. Следовательно, в качестве оценки параметра λ возьмём величину $\lambda^*=\frac{1}{\bar{x}_в}$. Тогда теоретическое распределение будет иметь вид

$$f(x)=\begin{cases}0,\ x<0;\\ \frac{1}{\bar{x}_в}e^{-\frac{1}{\bar{x}_в}x}, x\geq 0.\end{cases}$$

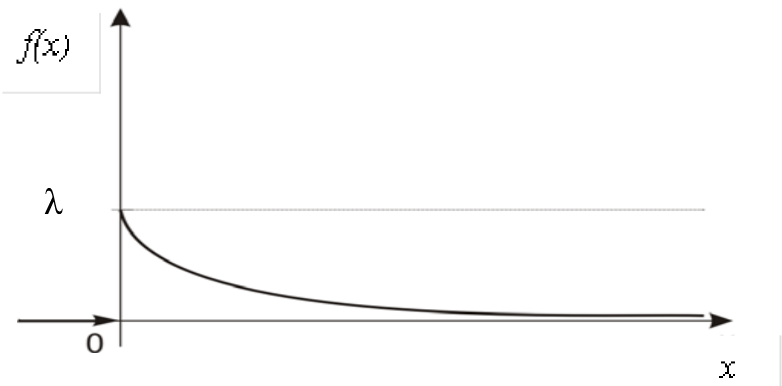

Рис. 3.5. График функции плотности равномерного распределения

Вероятности p_i вычисляются по формуле

$$p_i=p_i\left(x_i<X<x_{i+1}\right)=\int_{x_i}^{x_{i+1}}f\left(x\right)dx=\int_{x_i}^{x_{i+1}}\frac{1}{\bar{x}_в}e^{-\frac{1}{\bar{x}_в}x}dx=e^{-\frac{1}{\bar{x}_в}x_i}-e^{-\frac{1}{\bar{x}_в}x_{i+1}}.$$

Чаще всего пользуются приближенной формулой

$$p_i=p_i\left(x_i<X<x_{i+1}\right)=\int_{x_i}^{x_{i+1}}f\left(x\right)dx=\frac{1}{\bar{x}_в}e^{-\frac{1}{\bar{x}_в}\tilde{x}_i}\cdot\Delta,$$

где $\tilde{x}_i=\frac{x_i+x_{i+1}}{2}$.

Тогда теоретические частоты вычисляются по формуле

$$n_i=p_i\cdot n,$$

где n – объем выборки.

§6. Интервальные оценки математического ожидания и среднего квадратического отклонения

При выборке малого объема точечная оценка может значительно отличаться от оцениваемого параметра, что приводит к грубым ошибкам. Поэтому в таком случае лучше пользоваться *интервальными оценками*, то есть указывать интервал, в который с заданной вероятностью попадает истинное значение оцениваемого параметра. Разумеется, чем меньше длина этого интервала, тем точнее оценка параметра.

Поэтому, если для оценки θ^* некоторого параметра θ справедливо неравенство $|\ \theta^* - \theta\ | < \delta$, число $\delta > 0$ характеризует *точность оценки* (чем меньше δ, тем точнее оценка). Но статистические методы позволяют говорить только о том, что это неравенство выполняется с некоторой вероятностью.

Определение. *Доверительным интервалом* называется интервал $\theta^* - \delta < \theta < \theta^* + \delta$, в котором с заданной вероятностью γ заключено истинное значение неизвестного параметра θ. Величина γ называется доверительной вероятностью или надежностью. Из определения доверительной вероятности следует, что

$$\gamma = P\ (\ \theta^* - \delta < \theta < \theta^* + \delta\ ).$$

На практике обычно берут γ=0,95 или γ=0,99.

Рассмотрим построение доверительных интервалов для математического ожидания и среднего квадратического отклонения нормально распределенной случайной величины.

1. *Доверительный интервал для оценки математического ожидания нормального распределения при неизвестной дисперсии.*

Если известно, что исследуемая случайная величина X распределена по нормальному закону с неизвестным средним квадратическим отклонением, то для поиска доверительного интервала для ее математического ожидания a построим новую случайную величину

$$T = \frac{\bar{x}_в - a}{s} \cdot \sqrt{n},$$

где $\bar{x}_в$ – выборочное среднее; s^2 – исправленная дисперсия; n – объем выборки. Эта случайная величина, возможные значения которой будем обозначать через t, имеет распределение Стьюдента с $k = n - 1$ степенями свободы. Плотность распределения Стьюдента явным образом не зависит от a и σ (неизвестное СКО), а зависит только от n.

По определению доверительного интервала с заданной надежностью γ имеем

$$\gamma = P\left(\bar{x}_в - \delta < a < \bar{x}_в + \delta\right) = P\left(\left|\bar{x}_в - a\right| < \delta\right) =$$
$$= P\left(\frac{\left|\bar{x}_в - a\right| \cdot \sqrt{n}}{s} < \frac{\delta \cdot \sqrt{n}}{s}\right) = P\left(\left|T\right| < t_\gamma\right),$$

где $t_\gamma = \dfrac{\delta\sqrt{n}}{s}$.

Число $t_\gamma(\kappa)$называется критерием Стьюдента для уровня значимости α и числа степеней свободы $k = n - 1$. Его определяем по таблице для распределения Стьюдента (прил. 3). При определении доверительных интервалов задаются обычно надежностями γ=1–α, равными 0,9; 0,95; 0,99. Затем из (3.6) определяем величину $\delta = \dfrac{s \cdot t_\gamma}{\sqrt{n}}$, которую находим из соотношения $t_\gamma = \dfrac{\delta\sqrt{n}}{s}$. Таким образом, доверительный интервал с надежностью γ для математического ожидания a при неизвестном s есть интервал вида

$$\bar{x}_в - \frac{t_\gamma \cdot s}{\sqrt{n}} < a < \bar{x}_в + \frac{t_\gamma \cdot s}{\sqrt{n}}.$$

При больших n (n >30) распределение Стьюдента практически совпадает с нормальным. В этом случае t_γ можно найти из уравнения

$$2\Phi(t_\gamma) = 1 - \alpha = \gamma.$$

Пример 1. По 25-ти деталям выборочные характеристики прочности X составили: $x_в$ = 3; s = 1,5. Найти доверительный интервал для математического ожидания нормально распределенной случайной величины X и точность оценки δ при γ= 0,99.

Решение. Из таблицы распределения Стьюдента (см. прил. 3) находим, что t_γ (n = 25; $к$=25–1; α = 1–γ=1–0,99=0,01) = 2,80.

Тогда

$$3 - \frac{2{,}8 \cdot 1{,}5}{\sqrt{25}} < a < 3 + \frac{2{,}8 \cdot 1{,}5}{\sqrt{25}},$$

или 2,16< a < 3,84 – доверительный интервал, в который попадает a с вероятностью 0,99. Точность оценки δ = 0,84.

2. *Доверительные интервалы для оценки среднего квадратического отклонения нормального распределения.*

Пусть по выборке объёма n получено исправленное среднее квадратическое отклонение $s = \sqrt{\dfrac{n}{n-1} D_в}$, которое является точечной оценкой среднего квадратического отклонения σ случайной величины X. Будем искать для s нормально распределенной случайной величины доверительный интервал вида $(s - \delta; s + \delta)$, где s – исправленное выборочное среднее квадратическое отклонение. По определению, доверительный интервал с заданной надежностью $P(|\sigma - s| < \delta) = \gamma$ имеет вид $s - \delta < \sigma < s + \delta$.

Запишем это неравенство в виде $s\left(1 - \dfrac{\delta}{s}\right) < \sigma < s\left(1 + \dfrac{\delta}{s}\right)$ или обозначим $\dfrac{\delta}{s} = q$ и, подставив q в рассмотренное выше неравенство, получим

$$s(1-q) < \sigma < s(1+q) \quad \text{при } q<1;$$
$$0 < \sigma < s(1+q) \quad \text{при } q>1,$$

где $q = q(n, \gamma)$.

Существуют таблицы, из которых можно найти q по заданным n и γ (прил. 5). Случайная величина q имеет распределение, зависящее только от n.

Для дисперсии оценка имеет вид:

$$s^2(1-q)^2 < D < s^2(1+q)^2 \text{ при } q<1;$$
$$0 < D < s^2(1+q)^2 \qquad \text{при } q>1.$$

Пример 2. Произведено 20 измерений одним прибором некоторой случайной величины, имеющей нормальное распределение. Исправленное выборочное среднее квадратическое отклонение случайных ошибок измерений равно 1,3. Найти точность прибора с надежностью 0,95.

Решение. Точность прибора – это среднее квадратическое отклонение σ случайных ошибок измерений. По условию задачи, n = 20; s = =1,3. Найдем доверительный интервал для σ при заданной надежности γ = 0,95. По прил. 5 находим q (n = 20; γ = 0,95) = 0,37. Следовательно, границы доверительного интервала: 1,3(1–0,37) = 0,819 и 1,3(1+0,37) = 1,781. Итак, 0,819 < σ < 1,781 с вероятностью 0,95.

3. *Доверительные интервалы для оценки дисперсии.*

Считаем, что, вообще говоря, математическое ожидание неизвестно, а известна только точечная несмещенная оценка дисперсии s^2. Тогда доверительный интервал для дисперсии, соответствующий доверительной вероятности $\gamma = 1-\alpha$, имеет вид

$$\frac{(n-1)s^2}{\chi_2^2} \le D \le \frac{(n-1)s^2}{\chi_1^2}. \qquad (3.7)$$

Так как по заданной вероятности γ можно построить множество доверительных интервалов для дисперсии, то принято χ_1^2 и χ_2^2 выбирать так, чтобы

$$P(\chi^2 > \chi_1^2) = 1-\frac{\alpha}{2}, \text{ а } P(\chi^2 < \chi_2^2) = \frac{\alpha}{2},$$

где числа χ_1^2 и χ_2^2 находят по прил. 4 при числе степеней свободы

$k = n-1$ и соответственно при $p = 1-\dfrac{\alpha}{2}$ и $p = \dfrac{\alpha}{2}$.

Пример 3. Пользуясь 90%-ным доверительным интервалом, оцените в условиях примера 2 изменение прочности деталей во всей генеральной совокупности.

Решение. По условию, n=25; s=1,5; γ=0,9. Найдем доверительный интервал для оценки дисперсии.

Согласно формуле (3.7) $\sqrt{\dfrac{(n-1)s^2}{\chi_2^2}} \le \sigma \le \sqrt{\dfrac{(n-1)s^2}{\chi_1^2}}$, а так как при k=n–1=25–1=24 верхняя доверительная граница равна

$$\chi_2^2(p;\kappa) = \chi_2^2\left(\frac{1-\gamma}{2};\kappa\right) = \chi_2^2(0{,}05;24) = 36{,}4,$$

нижняя определяется как

$$\chi_1^2(p;\kappa)=\chi_1^2\left(1-\frac{1-\gamma}{2};\kappa\right)=\chi_1^2(0{,}95;24)=13{,}8\,\text{(см. прил.4)},$$

то $\sqrt{\dfrac{24\cdot 1{,}5^2}{36{,}4}}\le\sigma\le\sqrt{\dfrac{24\cdot 1{,}5^2}{13{,}8}}$ или $1{,}22\le\sigma\le 1{,}98$ – эта оценка не симметрична относительно σ.