Тема 4. Доверительные интервалы и границы.

1. Доверительный интервал, нижняя и верхняя доверительные границы.

Всякая точечная оценка сообщает лишь одно значение, которое принимается за приближенное значение оцениваемой величины, при этом полученное значение в большинстве случаев, конечно, не совпадает с истинным значением оцениваемой величины, поэтому в ряде случаев требуется указать интервал, в котором с большой вероятностью находится оцениваемая величина.

Пусть $(\xi_1,..., \xi_n)$ – серия наблюдений, θ – неизвестный скалярный параметр и Θ – множество допустимых значений параметра θ.

Определение 4.1.

Пусть $T_1(\xi_1,..., \xi_n)$ и $T_2(\xi_1,..., \xi_n)$ – статистики. Интервал $(T_1(\xi_1,..., \xi_n); T_2(\xi_1,..., \xi_n))$ называется *доверительным интервалом для величины* $\tau(\theta)$ *с уровнем доверия (доверительной вероятностью)* $P_д$ ($0 < P_д < 1$), если:

$$\inf_{\theta\in\Theta} P\{T_1(\xi_1,..., \xi_n) < \tau(\theta) < T_2(\xi_1,..., \xi_n)\} = P_д .$$

Из условия 2) определения 4.1 следует, что статистики $T_1(\xi_1,..., \xi_n)$ и $T_2(\xi_1,..., \xi_n)$ устроены таким образом, что каким бы ни оказалось значение параметра θ величина $\tau(\theta)$ «накрывается» интервалом $(T_1(\xi_1,..., \xi_n); T_2(\xi_1,..., \xi_n))$ с вероятностью не меньше чем $P_д$.

Определение 4.2.

Статистика $T(\xi_1,..., \xi_n)$ называется *верхней доверительной границей для* $\tau(\theta)$ *с уровнем доверия (доверительной вероятностью)* $P_д$ ($0 < P_д < 1$), если:

$$\inf_{\theta\in\Theta} P\{\tau(\theta) < T(\xi_1,..., \xi_n)\} = P_д .$$

Определение 4.3.

Статистика $T(\xi_1,..., \xi_n)$ называется *нижней доверительной границей для* $\tau(\theta)$ *с уровнем доверия (доверительной вероятностью)* $P_д$ *(*$0 < P_д < 1$*)*, если:

$$\inf_{\theta\in\Theta} P\{T(\xi_1,..., \xi_n) < \tau(\theta)\} = P_д .$$

Существует достаточно общий метод построения доверительных интервалов и границ, который основывается на функции специального вида, называемой центральной статистикой.

Определение 4.4.

Пусть $(\xi_1,..., \xi_n)$ – совокупность наблюдений и случайная функция $\varphi(\xi_1,..., \xi_n; \tau(\theta))$ зависит как от величин $(\xi_1,..., \xi_n)$, так и от неизвестной величины $\tau(\theta)$. Случайная функция $\varphi(\xi_1,..., \xi_n; \tau(\theta))$ называется *центральной статистикой для величины* $\tau(\theta)$, если:

1) распределение $\varphi(\xi_1,..., \xi_n; \tau(\theta))$ известно (то есть никаким образом не зависит от неизвестного параметра θ),

2) при всех реализациях $(x_1,..., x_n)$ серии наблюдений $(\xi_1,..., \xi_n)$ одновременно функция $\varphi(x_1,..., x_n; t)$ непрерывна и строго монотонна по t (например, при всех $(x_1,..., x_n)$ функция $\varphi(x_1,..., x_n; t)$ непрерывна и возрастает по t).

Предположим, что некоторым образом построена центральная статистика для $\tau(\theta)$ – $\varphi(\xi_1,..., \xi_n; \tau(\theta))$, поскольку функция распределения $\varphi(\xi_1,..., \xi_n; \tau(\theta))$ известна (условие 1), то всегда можно найти числа y_1^* и y_2^* такие, что:

$$\begin{cases} P\{y_1^* < \varphi(\xi_1,\dots,\xi_n;\tau(\theta)) < y_2^*\} = P_\partial \\ y_2^* - y_1^* = \min\limits_{y_1,y_2:P\{y_1<\varphi<y_2\}=P_\partial} (y_2 - y_1) \end{cases}.$$

Поскольку функция $\varphi(x_1,\dots, x_n;t)$ непрерывна по t при всех реализациях $(x_1,\dots, x_n)$, то при каждом $(x_1,\dots, x_n)$ существуют решения t_1^* и t_2^* системы уравнений (рисунок 4.1):

$$\begin{cases} y_1^* = \varphi(x_1,\dots,x_n;t_1^*) \\ y_2^* = \varphi(x_1,\dots,x_n;t_2^*) \end{cases} \Leftrightarrow \begin{cases} t_1^* = t_1^*(x_1,\dots,x_n;y_1^*) \\ t_2^* = t_2^*(x_1,\dots,x_n;y_2^*) \end{cases}$$

Рисунок 4.1.

Если функция $\varphi(x_1,\dots, x_n;t)$ возрастает по t при всех реализациях, тогда события $\{y_1^* < \varphi(\xi_1,\dots,\xi_n;\tau(\theta)) < y_2^*\}$ и $\{t_1^*(\xi_1,\dots,\xi_n;y_1^*) < \tau(\theta) < t_2^*(\xi_1,\dots,\xi_n;y_2^*)\}$ эквивалентны и вероятности событий равны, то есть:

$$P\{t_1^*(\xi_1,\dots,\xi_n;y_1^*) < \tau(\theta) < t_2^*(\xi_1,\dots,\xi_n;y_2^*)\} = P\{y_1^* < \varphi(\xi_1,\dots,\xi_n;\tau(\theta)) < y_2^*\} = P_\partial .$$

Пусть статистики $T_1(\xi_1,\dots,\xi_n) = t_1^*(\xi_1,\dots,\xi_n;y_1^*)$ и $T_2(\xi_1,\dots,\xi_n) = t_2^*(\xi_1,\dots,\xi_n;y_2^*)$, тогда интервал $(T_1(\xi_1,\dots, \xi_n);T_2(\xi_1,\dots, \xi_n))$ является доверительным интервалом для $\tau(\theta)$ с уровнем доверия P_∂, поскольку для всех допустимых значений параметра θ:

$$P\{T_1(\xi_1,\dots, \xi_n) < \tau(\theta) < T_2(\xi_1,\dots, \xi_n)\} = P_\partial ,$$

следовательно,

$$\inf_{\theta\in\Theta} P\{T_1(\xi_1,\dots,\xi_n) < \tau(\theta) < T_2(\xi_1,\dots,\xi_n)\} = \inf_{\theta\in\Theta} P_\partial = P_\partial .$$

Если функция $\varphi(x_1,\dots, x_n;t)$ убывает по t при всех реализациях, тогда эквивалентны события $\{y_1^* < \varphi(\xi_1,\dots,\xi_n;\tau(\theta)) < y_2^*\}$ и $\{t_2^*(\xi_1,\dots,\xi_n;y_2^*) < \tau(\theta) < t_1^*(\xi_1,\dots,\xi_n;y_1^*)\}$ и равны вероятности:

$$P\{t_2^*(\xi_1,\dots,\xi_n;y_2^*) < \tau(\theta) < t_1^*(\xi_1,\dots,\xi_n;y_1^*)\} = P\{y_1^* < \varphi(\xi_1,\dots,\xi_n;\tau(\theta)) < y_2^*\} = P_\partial .$$

Пусть статистики $T_1(\xi_1,\dots,\xi_n) = t_2^*(\xi_1,\dots,\xi_n;y_2^*)$ и $T_2(\xi_1,\dots,\xi_n) = t_1^*(\xi_1,\dots,\xi_n;y_1^*)$, тогда интервал $(T_1(\xi_1,\dots, \xi_n);T_2(\xi_1,\dots, \xi_n))$ является доверительным интервалом для $\tau(\theta)$ с уровнем доверия P_∂, поскольку для всех допустимых значений параметра θ:

$$P\{T_1(\xi_1,\dots, \xi_n) < \tau(\theta) < T_2(\xi_1,\dots, \xi_n)\} = P_\partial ,$$

тогда,

$$\inf_{\theta\in\Theta} P\{T_1(\xi_1,\dots,\xi_n) < \tau(\theta) < T_2(\xi_1,\dots,\xi_n)\} = \inf_{\theta\in\Theta} P_\partial = P_\partial$$

Аналогичным образом, с помощью центральной статистики $\varphi(\xi_1,\dots, \xi_n;\tau(\theta))$ могут быть построены доверительные границы.

2. Доверительный интервал для математического ожидания нормального распределения с известной дисперсией.

Пусть $(\xi_1,\dots, \xi_n)$ – выборка из нормального распределения с неизвестным математическим ожиданием m и известной дисперсией σ^2, построим доверительный интервал для математического ожидания m с уровнем доверия P_∂.

Величины ξ_i выборки имеют нормальное распределение и независимы, поэтому совокупность величин $(\xi_1,\dots, \xi_n)$ имеет многомерное нормальное распределение, и статистика $\frac{1}{n}\sum_{i=1}^{n}\xi_i$ также имеет нормальное распределение с параметрами:

$$M_m\left[\frac{1}{n}\sum_{i=1}^{n}\xi_i\right]=\frac{1}{n}\sum_{i=1}^{n}M_m[\xi_i]=\frac{1}{n}\sum_{i=1}^{n}m=m\ ,$$

$$D_m\left[\frac{1}{n}\sum_{i=1}^{n}\xi_i\right]=\frac{1}{n^2}\sum_{i=1}^{n}D_m[\xi_i]=\frac{\sigma^2}{n}\,.$$

Тогда статистика $\varphi(\xi_1,...,\ \xi_n;m)$:

$$\varphi(\xi_1,...,\xi_n;m)=\frac{\frac{1}{n}\sum_{i=1}^{n}\xi_i-m}{\frac{\sigma^2}{\sqrt{n}}},$$

имеет нормальное распределение $N(0,1)$ не зависящее от неизвестного параметра m и одновременно при всех реализациях $(\xi_1,...,\ \xi_n)$ функция $\varphi(\xi_1,...,\ \xi_n;m)$ как функция m является непрерывной и убывающей. Согласно определению $\varphi(\xi_1,...,\ \xi_n;m)$ – центральная статистика для m . Выберем числа y_1^* и y_2^* так, чтобы выполнялись равенства:

$$\begin{cases} P\{y_1^*<\varphi(\xi_1,...,\xi_n;m)<y_2^*\}=P_\partial \\ y_2^*-y_1^*=\min\limits_{y_1,y_2:P\{y_1<\varphi<y_2\}=P_\partial}(y_2-y_1) \end{cases}$$

или

$$\begin{cases} \Phi(y_2^*)-\Phi(y_1^*)=P_\partial \\ y_2^*-y_1^*=\min\limits_{y_1,y_2:P\{y_1<\varphi<y_2\}=P_\partial}(y_2-y_1) \end{cases},$$

где $\Phi(y)$ - функция распределения нормальной случайной величины $N(0,1)$. Для нахождения минимума функции y_2-y_1 при условии $\Phi(y_2)-\Phi(y_1)=P_\partial$ воспользуемся методом множителей Лагранжа, с функцией Лагранжа:

$$\Lambda(y_1,y_2,\lambda)=(y_2-y_1)-\lambda(\Phi(y_2)-\Phi(y_1)-P_\partial)\ ,$$

которая приводит к системе:

$$\begin{cases} \left.\frac{\partial\Lambda}{\partial y_1}\right|_{(y_1,y_2,\lambda)=(y_1^*,y_2^*,\lambda^*)}=0 \\ \left.\frac{\partial\Lambda}{\partial y_2}\right|_{(y_1,y_2,\lambda)=(y_1^*,y_2^*,\lambda^*)}=0 \\ \left.\frac{\partial\Lambda}{\partial \lambda}\right|_{(y_1,y_2,\lambda)=(y_1^*,y_2^*,\lambda^*)}=0 \end{cases}\Leftrightarrow$$

$$\begin{cases} -1+\lambda^*\Phi'(y_1^*)=0 \\ 1-\lambda^*\Phi'(y_2^*)=0 \\ \Phi(y_2^*)-\Phi(y_1^*)-P_\partial=0 \end{cases}\Leftrightarrow\begin{cases} \lambda^*=\frac{1}{\Phi'(y_1^*)} \\ 1-\frac{\Phi'(y_2^*)}{\Phi'(y_1^*)}=0 \\ \Phi(y_2^*)-\Phi(y_1^*)-P_\partial=0 \end{cases}\Leftrightarrow\begin{cases} \lambda^*=\frac{1}{\Phi'(y_1^*)} \\ \Phi'(y_2^*)=\Phi'(y_1^*) \\ \Phi(y_2^*)-\Phi(y_1^*)-P_\partial=0 \end{cases}\Leftrightarrow$$

$$\begin{cases} \lambda^* = \dfrac{1}{\Phi'(y_1^*)} \\ \dfrac{1}{\sqrt{2\pi}} e^{-\frac{(y_2^*)^2}{2}} = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{(y_1^*)^2}{2}} \\ \Phi(y_2^*) - \Phi(y_1^*) - P_д = 0 \end{cases} \Leftrightarrow \begin{cases} \lambda^* = \dfrac{1}{\Phi'(y_1^*)} \\ (y_2^*)^2 = (y_1^*)^2 \\ \Phi(y_2^*) - \Phi(y_1^*) - P_д = 0 \end{cases}.$$

У второго уравнения системы $(y_2^*)^2 = (y_1^*)^2$ или $(y_2^*)^2 - (y_1^*)^2 = (y_2^* - y_1^*)(y_2^* + y_1^*) = 0$, очевидно, имеется только два решения $y_2^* = y_1^*$ и $y_2^* = -y_1^*$, первое решение не удовлетворяет третьему уравнению системы $\Phi(y_2^*) - \Phi(y_1^*) - P_д = 0$, тогда:

$$\begin{cases} \lambda^* = \dfrac{1}{\Phi'(y_1^*)} \\ y_2^* = -y_1^* \\ \Phi(-y_1^*) - \Phi(y_1^*) - P_д = 0 \end{cases}$$

Используя свойство функции нормального распределения $\Phi(x) = 1 - \Phi(-x)$ получим:

$$\begin{cases} \lambda = \dfrac{1}{\Phi'(y_1^*)} \\ y_2^* = -y_1^* \\ 1 - \Phi(y_1^*) - \Phi(y_1^*) - P_д = 0 \end{cases} \Leftrightarrow \begin{cases} \lambda = \dfrac{1}{\Phi'(y_1^*)} \\ y_2^* = -y_1^* \\ \Phi(y_1^*) = \dfrac{1 - P_д}{2} \end{cases}.$$

Таким образом, y_1^* есть квантиль уровня $\dfrac{1 - P_д}{2}$ распределения $N(0,1)$ и $y_2^* = -y_1^*$. Значению y_2^* можно придать иную интерпретацию:

$$\begin{cases} \Phi(y_2^*) = 1 - \Phi(-y_2^*) \\ y_2^* = -y_1^* \\ \Phi(y_1^*) = \dfrac{1 - P_д}{2} \end{cases} \Rightarrow \Phi(y_2^*) = 1 - \Phi(-y_2^*) = 1 - \Phi(y_1^*) = 1 - \frac{1 - P_д}{2} = \frac{1 + P_д}{2},$$

то есть y_2^* является квантилью уровня $\dfrac{1 + P_д}{2}$ распределения $N(0,1)$. Таким образом, получим равенство для вероятностей:

$$P\{y_1^* < \varphi(\xi_1, ..., \xi_n; m) < y_2^*\} = P_д.$$

Преобразовывая неравенства, получим:

$$P\left\{ y_1^* < \frac{\frac{1}{n}\sum_{i=1}^{n} \xi_i - m}{\frac{\sigma}{\sqrt{n}}} < y_2^* \right\} = P_д,$$

$$P\left\{ y_1^* \frac{\sigma}{\sqrt{n}} < \frac{1}{n}\sum_{i=1}^{n} \xi_i - m < y_2^* \frac{\sigma}{\sqrt{n}} \right\} = P_д$$

$$P\left\{ -\frac{1}{n}\sum_{i=1}^{n} \xi_i + y_1^* \frac{\sigma}{\sqrt{n}} < -m < -\frac{1}{n}\sum_{i=1}^{n} \xi_i + y_2^* \frac{\sigma}{\sqrt{n}} \right\} = P_д$$

$$P\left\{\frac{1}{n}\sum_{i=1}^{n}\xi_i - y_2^*\frac{\sigma}{\sqrt{n}} < m < \frac{1}{n}\sum_{i=1}^{n}\xi_i - y_1^*\frac{\sigma}{\sqrt{n}}\right\} = P_д .$$

Преобразование неравенств фактически является нахождением решения системы:

$$\begin{cases} y_1^* = \varphi(x_1,..., x_n; m_1^*) \\ y_2^* = \varphi(x_1,..., x_n; m_2^*) \end{cases} \leftrightarrow \begin{cases} y_1^* = \dfrac{\frac{1}{n}\sum_{i=1}^{n} x_i - m_1^*}{\frac{\sigma}{\sqrt{n}}} \\ y_2^* = \dfrac{\frac{1}{n}\sum_{i=1}^{n} x_i - m_2^*}{\frac{\sigma}{\sqrt{n}}} \end{cases} \Leftrightarrow \begin{cases} m_1^* = \frac{1}{n}\sum_{i=1}^{n} x_i - y_1^*\frac{\sigma}{\sqrt{n}} \\ m_2^* = \frac{1}{n}\sum_{i=1}^{n} x_i - y_2^*\frac{\sigma}{\sqrt{n}} = \end{cases} \leftrightarrow \begin{cases} m_1^* = m_1^*(x_1,..., x_n; y_1^*) \\ m_2^* = m_2^*(x_1,..., x_n; y_2^*) \end{cases}.$$

Таким образом, при всяком значении параметра m :

$$P\left\{\frac{1}{n}\sum_{i=1}^{n}\xi_i - y_2^*\frac{\sigma}{\sqrt{n}} < m < \frac{1}{n}\sum_{i=1}^{n}\xi_i - y_1^*\frac{\sigma}{\sqrt{n}}\right\} = P_д ,$$

тогда интервал ($y_2^* = -y_1^*$):

$$\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i - y_2^*\frac{\sigma}{\sqrt{n}}; \frac{1}{n}\sum_{i=1}^{n}\xi_i + y_2^*\frac{\sigma}{\sqrt{n}}\right),$$

где y_2^* – является квантилью уровня $\frac{1+P_д}{2}$ распределения $N(0,1)$, является доверительным интервалом для m с уровнем доверия $P_д$.

3. Доверительный интервал для дисперсии нормального распределения с известным математическим ожиданием.

Пусть $(\xi_1,..., \xi_n)$ – выборка из нормального распределения с известным математическим ожиданием m и неизвестной дисперсией σ^2, построим доверительный интервал для дисперсии σ^2 с уровнем доверия $P_д$.

Рассмотрим статистику $\varphi(\xi_1,..., \xi_n; \sigma^2)$:

$$\varphi(\xi_1,..., \xi_n; \sigma^2) = \frac{n}{\sigma^2}\cdot\frac{1}{n}\sum_{i=1}^{n}(\xi_i - m)^2 = \sum_{i=1}^{n}\left(\frac{\xi_i - m}{\sigma}\right)^2 .$$

Поскольку случайные величины $\frac{\xi_i - m}{\sigma}$ имеют нормальное распределение $N(0,1)$ и независимы, то статистика $\varphi(\xi_1,..., \xi_n; \sigma^2)$ имеет распределение $\chi^2(n)$ («хи-квадрат с n степенями свободы») и кроме того одновременно при всех реализациях выборки $(\xi_1,..., \xi_n)$ функция $\varphi(\xi_1,..., \xi_n; \sigma^2)$ как функция параметра σ^2 :

$$\varphi(\xi_1,..., \xi_n; \sigma^2) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(\xi_i - m)^2$$

является непрерывной и убывающей. Таким образом, статистика $\varphi(\xi_1,..., \xi_n; \sigma^2)$ является центральной статистикой для σ^2 .

Для построения доверительного интервала выберем числа y_1 и y_2 так, чтобы выполнялось равенство:

$$P\{y_1 < \varphi(\xi_1,..., \xi_n; \sigma^2) < y_2\} = P_д .$$

Для выполнения равенства достаточно, например, в качестве y_1 взять квантиль уровня $\frac{1-P_д}{2}$ распределения $\chi^2(n)$, а качестве y_2 – квантиль уровня $\frac{1+P_д}{2}$ распределения $\chi^2(n)$, действительно:

$$P\{y_1 < \varphi(\xi_1,..., \xi_n;\sigma^2) < y_2\} = P\{y_1 < \chi_n^2 < y_2\} = F_{\chi_n^2}(y_2) - F_{\chi_n^2}(y_1) = \frac{1+P_д}{2} - \frac{1-P_д}{2} = P_д,$$

где χ_n^2 – случайная величина, имеющая распределение $\chi^2(n)$, и $F_{\chi_n^2}(y)$ – функция распределения χ_n^2.

При таких значениях y_1 и y_2 получается так называемый «центральный интервал» (название обусловлено тем, что слева от y_1 «сосредоточена» вероятность $\frac{1-P_д}{2}$ и справа от y_2 «сосредоточена» вероятность $1-\frac{1+P_д}{2}=\frac{1-P_д}{2}$). Построение «наикратчайшего» доверительного интервала, то есть нахождение чисел y_1^* и y_2^* с наименьшей разностью $y_2 - y_1$ среди всех пар y_1 и y_2, удовлетворяющих $P\{y_1 < \varphi(\xi_1,..., \xi_n;\sigma^2) < y_2\} = P_д$, в данном случае является технически сложным ([Ивченко, Медведев 1] стр. 86), поэтому на практике ограничиваются более простым «центральным интервалом».

Преобразование неравенств приводит к следующему доверительному интервалу:

$$P\left\{ y_1 < \frac{1}{\sigma^2}\sum_{i=1}^{n}(\xi_i - m)^2 < y_2 \right\} = P_д,$$

$$P\left\{ \frac{y_1}{\sum_{i=1}^{n}(\xi_i - m)^2} < \frac{1}{\sigma^2} < \frac{y_2}{\sum_{i=1}^{n}(\xi_i - m)^2} \right\} = P_д,$$

$$P\left\{ \frac{\sum_{i=1}^{n}(\xi_i - m)^2}{y_2} < \sigma^2 < \frac{\sum_{i=1}^{n}(\xi_i - m)^2}{y_1} \right\} = P_д.$$

Поскольку последнее равенство справедливо при всяком значении σ^2, то интервал:

$$\left(\frac{\sum_{i=1}^{n}(\xi_i - m)^2}{y_2}; \frac{\sum_{i=1}^{n}(\xi_i - m)^2}{y_1} \right),$$

где y_1 и y_2 – квантили уровней $\frac{1-P_д}{2}$ и $\frac{1+P_д}{2}$ распределения $\chi^2(n)$ соответственно, является доверительным интервалом для σ^2 с уровнем доверия $P_д$.

Нетрудно также получить и доверительный интервал для с.к.о. σ, действительно, поскольку:

$$\begin{cases} \dfrac{\sum_{i=1}^{n}(\xi_i - m)^2}{y_2} < \sigma^2 < \dfrac{\sum_{i=1}^{n}(\xi_i - m)^2}{y_1} \\ \sigma > 0 \end{cases} \Leftrightarrow \sqrt{\frac{\sum_{i=1}^{n}(\xi_i - m)^2}{y_2}} < \sigma < \sqrt{\frac{\sum_{i=1}^{n}(\xi_i - m)^2}{y_1}},$$

то

$$P\left\{\sqrt{\frac{\sum_{i=1}^{n}(\xi_i-m)^2}{y_2}}<\sigma<\sqrt{\frac{\sum_{i=1}^{n}(\xi_i-m)^2}{y_1}}\right\}=P\left\{\frac{\sum_{i=1}^{n}(\xi_i-m)^2}{y_2}<\sigma^2<\frac{\sum_{i=1}^{n}(\xi_i-m)^2}{y_1}\right\}=P_д\,,$$

тогда при тех же значениях y_1 и y_2 интервал:

$$\left(\sqrt{\frac{\sum_{i=1}^{n}(\xi_i-m)^2}{y_2}};\sqrt{\frac{\sum_{i=1}^{n}(\xi_i-m)^2}{y_1}}\right)$$

является доверительным интервалом для σ с уровнем доверия $P_д$.

4. Доверительный интервал для дисперсии нормального распределения с неизвестным математическим ожиданием.

Теорема 4.5. (Фишер)

Пусть $(\xi_1,..., \xi_n)$ – выборка из нормального распределения $N(m,\sigma^2)$, статистики $\hat{m}_1(\xi_1,...,\xi_n)=\frac{1}{n}\sum_{i=1}^{n}\xi_i$ и $\tilde{\mu}_2(\xi_1,...,\xi_n)=\frac{1}{n-1}\sum_{i=1}^{n}(\xi_i-\hat{m}_1)^2$, тогда:

1) Статистика $\frac{(n-1)\tilde{\mu}_2(\xi_1,..., \xi_n)}{\sigma^2}$ имеет распределение $\chi^2(n-1)$;

2) Статистики $\hat{m}_1(\xi_1,..., \xi_n)$ и $\tilde{\mu}_2(\xi_1,..., \xi_n)$ – независимые случайные величины.

Доказательство:

1) Преобразуем статистику $\frac{(n-1)\tilde{\mu}_2(\xi_1,..., \xi_n)}{\sigma^2}$ следующим образом:

$$\frac{n-1}{\sigma^2}\tilde{\mu}_2(\xi_1,...,\xi_n)=\frac{n-1}{\sigma^2}\frac{1}{n-1}\sum_{i=1}^{n}(\xi_i-\hat{m}_1)^2=\sum_{i=1}^{n}\left(\frac{\xi_i-\hat{m}_1}{\sigma}\right)^2=\sum_{i=1}^{n}\left(\frac{\xi_i-m+m-\hat{m}_1}{\sigma}\right)^2=$$

$$=\sum_{i=1}^{n}\left(\frac{\xi_i-m}{\sigma}+\frac{m-\hat{m}_1}{\sigma}\right)^2=\sum_{i=1}^{n}\left(\left(\frac{\xi_i-m}{\sigma}\right)^2+2\left(\frac{\xi_i-m}{\sigma}\right)\left(\frac{m-\hat{m}_1}{\sigma}\right)+\left(\frac{m-\hat{m}_1}{\sigma}\right)^2\right)=$$

$$=\sum_{i=1}^{n}\left(\frac{\xi_i-m}{\sigma}\right)^2+2\left(\frac{m-\hat{m}_1}{\sigma}\right)\sum_{i=1}^{n}\left(\frac{\xi_i-m}{\sigma}\right)+\sum_{i=1}^{n}\left(\frac{\hat{m}_1-m}{\sigma}\right)^2=$$

$$=\sum_{i=1}^{n}\left(\frac{\xi_i-m}{\sigma}\right)^2+2\left(\frac{m-\hat{m}_1}{\sigma}\right)\frac{1}{\sigma}\left(\sum_{i=1}^{n}\xi_i-\sum_{i=1}^{n}m\right)+n\left(\frac{\hat{m}_1-m}{\sigma}\right)^2=$$

$$=\sum_{i=1}^{n}\left(\frac{\xi_i-m}{\sigma}\right)^2+2\left(\frac{m-\hat{m}_1}{\sigma}\right)\frac{1}{\sigma}\left(n\hat{m}_1-nm\right)+n\left(\frac{\hat{m}_1-m}{\sigma}\right)^2=$$

$$=\sum_{i=1}^{n}\left(\frac{\xi_i-m}{\sigma}\right)^2+2\left(\frac{m-\hat{m}_1}{\sigma}\right)n\left(\frac{\hat{m}_1-m}{\sigma}\right)+n\left(\frac{\hat{m}_1-m}{\sigma}\right)^2=$$

$$=\sum_{i=1}^{n}\left(\frac{\xi_i-m}{\sigma}\right)^2-2n\left(\frac{m-\hat{m}_1}{\sigma}\right)\left(\frac{m-\hat{m}_1}{\sigma}\right)+n\left(\frac{\hat{m}_1-m}{\sigma}\right)^2=$$

$$= \sum_{i=1}^{n}\left(\frac{\xi_i - m}{\sigma}\right)^2 - n\left(\frac{\hat{m}_1 - m}{\sigma}\right)^2 = \sum_{i=1}^{n}\left(\frac{\xi_i - m}{\sigma}\right)^2 - n\left(\frac{\frac{1}{n}\sum_{i=1}^{n}\xi_i - m}{\sigma}\right)^2 =$$

$$= \sum_{i=1}^{n}\left(\frac{\xi_i - m}{\sigma}\right)^2 - n\left(\frac{\frac{1}{n}\sum_{i=1}^{n}(\xi_i - m)}{\sigma}\right)^2 = \sum_{i=1}^{n}\left(\frac{\xi_i - m}{\sigma}\right)^2 - \frac{1}{n}\left(\sum_{i=1}^{n}\left(\frac{\xi_i - m}{\sigma}\right)\right)^2 .$$

Пусть $\xi = (\xi_1, ..., \xi_n)$ есть вектор-столбец исходной выборки, определим вектор-столбец случайных величин $\alpha = (\alpha_1, \alpha_2 ..., \alpha_n)$:

$$\alpha_i = \frac{\xi_i - m}{\sigma},$$

$$i = \overline{1, n}$$

или в матричной форме,

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ ... \\ \alpha_n \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma} & 0 & ... & 0 \\ 0 & \frac{1}{\sigma} & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & 0 & \frac{1}{\sigma} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ ... \\ \xi_n \end{pmatrix} - \begin{pmatrix} \frac{m}{\sigma} \\ \frac{m}{\sigma} \\ ... \\ \frac{m}{\sigma} \end{pmatrix},$$

$$\alpha = A_\alpha \xi - b_\alpha$$

Поскольку случайные величины ξ_i имеют совместное нормальное распределение, то случайные величины α_i также имеют совместное нормальное распределение (как линейное преобразование совместно нормальных случайных величин). Легко видеть, что математическое ожидание $M[\alpha]$ есть нулевой вектор $\overline{0}_n$:

$$M[\alpha] = M[A_\alpha \xi + b_\alpha] = A_\alpha M[\xi] - b_\alpha =$$

$$= \begin{pmatrix} \frac{1}{\sigma} & 0 & ... & 0 \\ 0 & \frac{1}{\sigma} & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & 0 & \frac{1}{\sigma} \end{pmatrix} \begin{pmatrix} M[\xi_1] \\ M[\xi_2] \\ ... \\ M[\xi_n] \end{pmatrix} - \begin{pmatrix} \frac{m}{\sigma} \\ \frac{m}{\sigma} \\ ... \\ \frac{m}{\sigma} \end{pmatrix} = \begin{pmatrix} \frac{M[\xi_1] - m}{\sigma} \\ \frac{M[\xi_2] - m}{\sigma} \\ ... \\ \frac{M[\xi_n] - m}{\sigma} \end{pmatrix} = \begin{pmatrix} \frac{m - m}{\sigma} \\ \frac{m - m}{\sigma} \\ ... \\ \frac{m - m}{\sigma} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ ... \\ 0 \end{pmatrix} = \overline{0}_n .$$

Поскольку исходный вектор ξ является выборкой, то дисперсионная матрица $D[\xi]$ исходного вектора ξ является диагональной матрицей:

$$\left\| D[\xi] \right\|_{i,j} = \operatorname{cov}(\xi_i - m, \xi_j - m) = \begin{cases} D[\xi_i] & , i = j \\ 0 & , i \neq j \end{cases} = \begin{cases} \sigma^2 & , i = j \\ 0 & , i \neq j \end{cases},$$

$$D[\xi] = \sigma^2 E_n ,$$

где E_n – единичная матрица размера $n \times n$. Дисперсионная матрица $D[\alpha]$ вектора α является единичной матрицей E_n (A_α^* – транспонированная матрица A_α):

$$D[\alpha] = D[A_\alpha \xi - b_\alpha] = A_\alpha^* D[\xi] A_\alpha = A_\alpha^* \sigma^2 E_n A_\alpha = \sigma^2 A_\alpha^* A_\alpha =$$

$$= \sigma^2 \begin{pmatrix} \frac{1}{\sigma^2} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma^2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \frac{1}{\sigma^2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 1 \end{pmatrix} = E_n ,$$

Легко видеть, что в результате преобразования исходная статистика преобразуется к следующему виду:

$$\frac{n-1}{\sigma^2}\tilde{\mu}_2(\xi_1,\dots,\xi_n) = \sum_{i=1}^{n}\left(\frac{\xi_i - m}{\sigma}\right)^2 - \frac{1}{n}\left(\sum_{i=1}^{n}\left(\frac{\xi_i - m}{\sigma}\right)\right)^2 = \sum_{i=1}^{n}\alpha_i^2 - \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\alpha_i\right)^2 .$$

Пусть A_β – ортогональная матрица (т.е. $A_\beta^{-1} = A_\beta^*$, где A_β^* – транспонированная матрица A_β), в которой все элементы первой строки равны $\frac{1}{\sqrt{n}}$:

$$A_\beta = \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \dots & \frac{1}{\sqrt{n}} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \qquad (4.1)$$

Определим вектор-столбец случайных величин $\beta = (\beta_1, \beta_2, \dots, \beta_n)$:

$$\beta = A_\beta \alpha ,$$

$$\begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_n \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \dots & \frac{1}{\sqrt{n}} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_n \end{pmatrix} .$$

Случайные величины $\beta_i = \sum_{j=1}^{n} \left\| A_\beta \right\|_{ij} \alpha_j$ имеют совместно нормальное распределение, поскольку компоненты α_j имеют совместно нормальное распределение. Математическое ожидание $M[\beta]$ есть нулевой вектор $\overline{0}_n$, действительно:

$$M[\beta] = M[A_\beta \alpha] = A_\beta M[\alpha] = A_\beta \overline{0}_n = \overline{0}_n ,$$

и дисперсионная матрица $D[\beta]$ есть единичная матрица E_n:

$$D[\beta] = D[A_\beta \alpha] = A_\beta^* D[\alpha] A_\beta = A_\beta^* E_n A_\beta = A_\beta^* A_\beta = A_\beta^{-1} A_\beta = E_n ,$$

поскольку A_β – ортогональная матрица ($A_\beta^{-1} = A_\beta^*$). Таким образом, случайные величины β_i являются некоррелированными, а поскольку β_i имеют совместно нормальное распределение, то, следовательно, случайные величины β_i независимы.

Легко проверить, что ортогональное преобразование $\beta = A_\beta \alpha$ не изменяет вторую норму векторов, то есть $\sum_{i=1}^{n} \beta_i^2 = \sum_{i=1}^{n} \alpha_i^2$, действительно:

$$\sum_{i=1}^{n} \beta_i^2 = \beta^* \beta = (A_\beta \alpha)^* A_\beta \alpha = \alpha^* A_\beta^* A_\beta \alpha = \alpha^* A_\beta^{-1} A_\beta \alpha = \alpha^* E_n \alpha = \alpha^* \alpha = \sum_{i=1}^{n} \alpha_i^2 .$$

Из определения матрицы A_β (4.1):

$$\beta_1 = \sum_{j=1}^{n} \left\| A_\beta \right\|_{1j} \alpha_j = \sum_{j=1}^{n} \frac{1}{\sqrt{n}} \alpha_j = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \alpha_j \tag{4.2}$$

Таким образом, исходная статистика преобразуется к виду:

$$\frac{n-1}{\sigma^2} \tilde{\mu}_2(\xi_1,\dots,\xi_n) = \sum_{i=1}^{n} \alpha_i^2 - \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \alpha_i \right)^2 = \sum_{i=1}^{n} \beta_i^2 - \beta_1^2 = \beta_2^2 + \dots + \beta_n^2 \tag{4.3}$$

где все величины β_i имеют нормальное распределение $N(0,1)$ и независимы, поэтому статистика $\frac{n-1}{\sigma^2} \tilde{\mu}_2(\xi_1,\dots,\xi_n)$ имеет распределение $\chi^2(n-1)$.

2) Из (4.2) следует:

$$\hat{m}_1(\xi_1,\dots,\xi_n) = \frac{1}{n} \sum_{i=1}^{n} \xi_i = \frac{1}{n} \sum_{i=1}^{n} (\sigma\alpha_i + m) = \frac{\sigma}{n} \sum_{i=1}^{n} \alpha_i + m = \frac{\sigma}{\sqrt{n}} \beta_1 + m$$

Из (4.3) следует:

$$\tilde{\mu}_2(\xi_1,\dots,\xi_n) = \frac{\sigma^2}{n-1} (\beta_2^2 + \dots + \beta_n^2).$$

Поскольку случайные величины β_i независимы, то следовательно независимы $\hat{m}_1(\xi_1,\dots,\xi_n)$ и $\tilde{\mu}_2(\xi_1,\dots,\xi_n)$.

Теорема доказана.

Теорема 4.5 позволяет построить доверительный интервал для дисперсии нормального распределения в случае, когда математическое ожидание неизвестно. Пусть $(\xi_1,\dots,\xi_n)$ – выборка из нормального распределения $N(m,\sigma^2)$, из теоремы 4.5 следует, что статистика $\varphi(\xi_1,\dots,\xi_n;\sigma^2)$:

$$\varphi(\xi_1,\dots,\xi_n;\sigma^2) = \frac{n-1}{\sigma^2} \tilde{\mu}_2(\xi_1,\dots,\xi_n) = \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(\xi_i - \frac{1}{n} \sum_{i=1}^{n} \xi_i \right)^2$$

имеет распределение $\chi^2(n-1)$, не зависящее от неизвестных параметров m и σ^2, и одновременно при всех реализациях выборки $(\xi_1,\dots,\xi_n)$ функция $\varphi(\xi_1,\dots,\xi_n;\sigma^2)$ как функция σ^2 является непрерывной и убывающей. Следовательно, статистика $\varphi(\xi_1,\dots,\xi_n;\sigma^2)$ является центральной статистикой для σ^2. Пусть y_1 и y_2 – квантили уровней $\frac{1-P_д}{2}$ и $\frac{1+P_д}{2}$ распределения $\chi^2(n-1)$, тогда:

$$P\{ y_1 < \varphi(\xi_1,\dots,\xi_n;\sigma^2) < y_2 \} = P_д,$$

$$P\left\{ y_1 < \frac{n-1}{\sigma^2} \tilde{\mu}_2(\xi_1,\dots,\xi_n) < y_2 \right\} = P_д,$$

$$P\left\{ \frac{(n-1)\tilde{\mu}_2(\xi_1,\dots,\xi_n)}{y_2} < \sigma^2 < \frac{(n-1)\tilde{\mu}_2(\xi_1,\dots,\xi_n)}{y_1} \right\} = P_д.$$

Таким образом, интервал

$$\left(\frac{\sum_{i=1}^{n} \left(\xi_i - \frac{1}{n} \sum_{i=1}^{n} \xi_i \right)^2}{y_2} ; \frac{\sum_{i=1}^{n} \left(\xi_i - \frac{1}{n} \sum_{i=1}^{n} \xi_i \right)^2}{y_1} \right),$$

где y_1 и y_2 являются квантилями уровней $\frac{1-P_д}{2}$ и $\frac{1+P_д}{2}$ распределения $\chi^2(n-1)$, является доверительным интервалом для дисперсии σ^2 с уровнем доверия $P_д$. Заметим, что при тех же значениях y_1 и y_2 интервал

$$\left(\sqrt{\frac{\sum_{i=1}^{n}\left(\xi_i-\frac{1}{n}\sum_{i=1}^{n}\xi_i\right)^2}{y_2}};\sqrt{\frac{\sum_{i=1}^{n}\left(\xi_i-\frac{1}{n}\sum_{i=1}^{n}\xi_i\right)^2}{y_1}}\right)$$

является доверительным интервалом для с.к.о. σ с уровнем доверия $P_д$.

5. Доверительный интервал для математического ожидания нормального распределения с неизвестной дисперсией.

Пусть $(\xi_1,..., \xi_n)$ – выборка из нормального распределения с неизвестным математическим ожиданием m и неизвестной дисперсией σ^2, и требуется построить доверительный интервал для m с уровнем доверия $P_д$.

Рассмотрим статистику $\varphi(\xi_1,..., \xi_n; m)$:

$$\varphi(\xi_1,...,\xi_n;m)=\frac{\dfrac{\hat{m}_1-m}{\dfrac{\sigma}{\sqrt{n}}}}{\sqrt{\dfrac{\dfrac{(n-1)\tilde{\mu}_2}{\sigma^2}}{n-1}}}=\frac{\hat{m}_1-m}{\sqrt{\tilde{\mu}_2}}\sqrt{n}\,, \tag{4.4}$$

$$\hat{m}_1=\frac{1}{n}\sum_{i=1}^{n}\xi_i\,,\ \tilde{\mu}_2=\frac{1}{n-1}\sum_{i=1}^{n}(\xi_i-\hat{m}_1)^2\,.$$

Заметим, что:

1) $\dfrac{\hat{m}_1-m}{\dfrac{\sigma}{\sqrt{n}}}\sim N(0,1)$, поскольку все величины ξ_i имеют нормальное распределение и независимы в совокупности;

2) $\dfrac{\hat{m}_1-m}{\dfrac{\sigma}{\sqrt{n}}}$ и $\dfrac{(n-1)\tilde{\mu}_2}{\sigma^2}$ независимы, поскольку в силу теоремы 4.5 статистики $\hat{m}_1$ и $\tilde{\mu}_2$ независимы;

3) $\dfrac{(n-1)\tilde{\mu}_2}{\sigma^2}$ имеет распределение $\chi^2(n-1)$ в силу теоремы 4.5.

Из 1)-3) следует, что статистика $\varphi(\xi_1,..., \xi_n; m)$ имеет распределение Стьюдента с $n-1$ степенью свободы $T(n-1)$. Кроме того, при всех реализациях выборки $(\xi_1,..., \xi_n)$ функция $\varphi(\xi_1,..., \xi_n; m)$ как функция m является непрерывной и убывающей, следовательно, случайная функция $\varphi(\xi_1,..., \xi_n; m)$ является центральной статистикой для m.

Пусть y_1 и y_2 – квантили уровней $\frac{1-P_д}{2}$ и $\frac{1+P_д}{2}$ распределения $T(n-1)$, тогда:

$$P\{y_1<\varphi(\xi_1,..., \xi_n; m)<y_2\}=P_д\,,$$

$$P\left\{ y_1 < \frac{\hat{m}_1 - m}{\sqrt{\tilde{\mu}_2}}\sqrt{n} < y_2 \right\} = P_д ,$$

$$P\left\{ \hat{m}_1 - y_2\sqrt{\frac{\tilde{\mu}_2}{n}} < m < \hat{m}_1 - y_1\sqrt{\frac{\tilde{\mu}_2}{n}} \right\} = P_д .$$

Поскольку функция плотности вероятности распределения Стьюдента $T(n-1)$ является симметричным относительно нуля, то для функции распределения $F_{T_{n-1}}(t)$ справедливо равенство:

$$F_{T_{n-1}}(t) = 1 - F_{T_{n-1}}(-t) .$$

Отсюда следует, что $-y_1 = y_2$, действительно:

$$F_{T_{n-1}}(-y_1) = 1 - F_{T_{n-1}}(y_1) = 1 - \frac{1-P_д}{2} = \frac{1+P_д}{2} = F_{T_{n-1}}(y_2) .$$

Таким образом,

$$P\left\{ \hat{m}_1 - y_2\sqrt{\frac{\tilde{\mu}_2}{n}} < m < \hat{m}_1 + y_2\sqrt{\frac{\tilde{\mu}_2}{n}} \right\} = P_д$$

и следовательно интервал,

$$\left(\hat{m}_1 - y_2\sqrt{\frac{\tilde{\mu}_2}{n}} ; \hat{m}_1 + y_2\sqrt{\frac{\tilde{\mu}_2}{n}} \right),$$

в котором $\hat{m}_1 = \frac{1}{n}\sum_{i=1}^{n}\xi_i$, $\tilde{\mu}_2 = \frac{1}{n-1}\sum_{i=1}^{n}(\xi_i - \hat{m}_1)^2$ и y_2 – квантиль уровня $\frac{1+P_д}{2}$ распределения Стьюдента с $n-1$ степенью свободы, является доверительным интервалом для m с уровнем доверия $P_д$.

6. Метод построения центральной статистики.

Пусть ξ – случайная величина с непрерывной и возрастающей по x функцией распределения $F_\xi(x\,|\,\theta)$, возможно зависящей от параметра θ. Рассмотрим случайную величину $\eta = F_\xi(\xi\,|\,\theta)$, легко видеть, что функция распределения $F_\eta(y\,|\,\theta)$ случайной величины η:

$$F_\eta(y\,|\,\theta) = P\{\eta < y\} = P\{F_\xi(\xi\,|\,\theta) < y\} = \begin{cases} 0 & , y < 0 \\ P\{\xi < F_\xi^{-1}(y\,|\,\theta)\} & ,0 \le y \le 1 \\ 1 & ,1 < y \end{cases},$$

где обратная функция $F_\xi^{-1}(y\,|\,\theta)$ существует, поскольку функция распределения $F_\xi(\xi\,|\,\theta)$ непрерывна и возрастает. Заметим, что если $0 \le y \le 1$, то:

$$P\{\xi < F_\xi^{-1}(y\,|\,\theta)\} = F_\xi(F_\xi^{-1}(y\,|\,\theta)) = y ,$$

таким образом,

$$F_\eta(y) = \begin{cases} 0 & , y < 0 \\ y & ,0 \le y \le 1 \\ 1 & ,1 < y \end{cases},$$

и следовательно случайная величина η имеет равномерное распределение $R[0,1]$, не зависящее от параметра θ.

Пусть $(\xi_1,..., \xi_n)$ – совокупность наблюдений и $T(\xi_1,..., \xi_n)$ – статистика, функция распределения которой $F_T(t\,|\,\theta)$ непрерывна и возрастает по t и кроме того известна полностью либо известна с точностью до значения параметра θ. Рассмотрим случайную величину $\varphi(\xi_1,..., \xi_n \mid \theta) = F_T(T(\xi_1,..., \xi_n) \mid \theta)$, согласно рассмотренному выше свойству функции распределения, случайная величина $\varphi(\xi_1,..., \xi_n \mid \theta)$ имеет равномерное распределение $R[0,1]$, не зависящее от параметра θ. Если при фиксированных $\xi_1,\dots,\xi_n$ функция $F_T(T(\xi_1,..., \xi_n) \mid \theta)$ как функция параметра θ является непрерывной и монотонной, тогда $\varphi(\xi_1,..., \xi_n \mid \theta)$ по определению является центральной статистикой.

Для построения "центрального" доверительного интервала достаточно вычислить число $\alpha = \frac{1 - P_д}{2}$, где $P_д$ – уровень доверия, $0.5 < P_д < 1$, тогда:

$$P\{\alpha < \varphi(\xi_1,\dots,\xi_n \mid \theta) < 1 - \alpha\} = F_\varphi(1 - \alpha) - F_\varphi(\alpha),$$

где $F_\varphi(x)$ – функция распределения $\varphi(\xi_1,..., \xi_n \mid \theta)$. Поскольку $\varphi(\xi_1,..., \xi_n \mid \theta)$ имеет равномерное распределение $R[0,1]$, то $F_\varphi(1 - \alpha) = 1 - \alpha$ и $F_\varphi(\alpha) = \alpha$, тогда:

$$P\{\alpha < \varphi(\xi_1,..., \xi_n \mid \theta) < 1 - \alpha\} = 1 - \alpha - \alpha = 1 - 2\alpha = P_д,$$

$$P\{\alpha < F_T(T(\xi_1,..., \xi_n) \mid \theta) < 1 - \alpha\} = P_д.$$

Разрешая неравенства $\alpha < F_T(T(\xi_1,..., \xi_n) \mid \theta) < 1 - \alpha$ относительно θ, получим доверительный интервал. Если функция $F_T(T(\xi_1,..., \xi_n) \mid \theta)$ возрастает по θ, тогда:

$$P\{F_\theta^{-1}(T(\xi_1,..., \xi_n), \alpha) < \theta < F_\theta^{-1}(T(\xi_1,..., \xi_n), 1 - \alpha)\} = P_д.$$

Если функция $F_T(T(\xi_1,..., \xi_n) \mid \theta)$ убывает по θ, тогда:

$$P\{F_\theta^{-1}(T(\xi_1,..., \xi_n), 1 - \alpha) < \theta < F_\theta^{-1}(T(\xi_1,..., \xi_n), \alpha)\} = P_д.$$

Если функция $F_T(T(\xi_1,..., \xi_n) \mid \theta)$ возрастает по θ, то для построения нижней доверительной границы достаточно взять $\alpha = 1 - P_д$ и рассмотреть вероятность:

$$P\{\alpha < \varphi(\xi_1,\dots,\xi_n \mid \theta)\} = 1 - F_\varphi(\alpha) = 1 - \alpha = 1 - (1 - P_д) = P_д,$$

$$P\{\alpha < F_T(T(\xi_1,..., \xi_n) \mid \theta)\} = P_д,$$

$$P\{F_\theta^{-1}(T(\xi_1,..., \xi_n), \alpha) < \theta\} = P_д.$$

Для построения верхней доверительной границы достаточно взять $\alpha = P_д$ и рассмотреть вероятность:

$$P\{\varphi(\xi_1,\dots,\xi_n \mid \theta) < \alpha\} = F_\varphi(\alpha) = \alpha = P_д,$$

$$P\{F_T(T(\xi_1,..., \xi_n) \mid \theta) < \alpha\} = P_д,$$

$$P\{\theta < F_\theta^{-1}(T(\xi_1,..., \xi_n), \alpha)\} = P_д.$$

Если функция $F_T(T(\xi_1,..., \xi_n) \mid \theta)$ убывает по θ, то для построения нижней доверительной границы следует взять $\alpha = P_д$ и рассмотреть вероятность:

$$P\{\varphi(\xi_1,\dots,\xi_n \mid \theta) < \alpha\} = F_\varphi(\alpha) = \alpha = P_д,$$

$$P\{F_T(T(\xi_1,..., \xi_n) \mid \theta) < \alpha\} = P_д,$$

$$P\{F_\theta^{-1}(T(\xi_1,..., \xi_n), \alpha) < \theta\} = P_д.$$

Для построения верхней доверительной границы следует взять $\alpha = 1 - P_д$ и рассмотреть вероятность:

$$P\{\alpha < \varphi(\xi_1,\dots,\xi_n \mid \theta)\} = 1 - F_\varphi(\alpha) = 1 - \alpha = 1 - (1 - P_д) = P_д,$$

$$P\{\alpha < F_T(T(\xi_1,..., \xi_n) \mid \theta)\} = P_д,$$

$$P\{\theta < F_\theta^{-1}(T(\xi_1,..., \xi_n), \alpha)\} = P_д.$$

7. Построение доверительных интервалов на основе асимптотической нормальности. Доверительный интервал для вероятности события.

Пусть $(\xi_1,..., \xi_n)$ – серия наблюдений и случайная величина $\varphi(\xi_1,..., \xi_n | \theta)$ имеет асимптотически (при $n \to \infty$) нормальное распределение $N(m(\theta), \sigma^2(\theta))$:

$$\varphi(\xi_1,..., \xi_n | \theta) \sim N(m(\theta), \sigma^2(\theta)) \text{, при } n \to \infty \text{;}$$

В силу асимптотической нормальности:

$$\frac{\varphi(\xi_1,..., \xi_n | \theta) - m(\theta)}{\sigma(\theta)} \sim N(0,1) \text{, при } n \to \infty \text{,}$$

тогда для y_2 при больших n справедливо приближенное равенство для вероятностей:

$$P\left\{- y_2 < \frac{\varphi(\xi_1,..., \xi_n | \theta) - m(\theta)}{\sigma(\theta)} < y_2\right\} \approx \Phi(y_2) - \Phi(-y_2) = \Phi(y_2) - (1 - \Phi(y_2)) = 2\Phi(y_2) - 1 .$$

Пусть y_2 является квантилью распределения $N(0,1)$ уровня $\frac{1 + P_д}{2}$, где $P_д$ – уровень доверия:

$$\Phi(y_2) = \frac{1 + P_д}{2} ,$$

тогда,

$$P\left\{- y_2 < \frac{\varphi(\xi_1,..., \xi_n | \theta) - m(\theta)}{\sigma(\theta)} < y_2\right\} \approx 2\Phi(y_2) - 1 = 2\frac{1 + P_д}{2} - 1 = P_д .$$

Разрешая неравенство слева относительно θ, получим «приближенный» доверительный интервал:

$$P\{T_1(\xi_1,..., \xi_n, y_2) < \theta < T_2(\xi_1,..., \xi_n, y_2)\} \approx P_д .$$

Воспользуемся вышеизложенным методом для построения «приближенного» доверительного интервала неизвестной вероятности события в схеме n независимых испытаний. Пусть $(\xi_1,..., \xi_n)$ – выборка, в которой каждая случайная величина ξ_i является бинарной и принимает значение 1 с некоторой неизвестной вероятностью p и значение 0 с вероятностью $1 - p$:

$$\xi_i = \begin{cases} 1 & , p \\ 0 & ,1 - p \end{cases} .$$

Требуется построить приближенный доверительный интервал для вероятности p. Рассмотрим случайную величину $\varphi(\xi_1,..., \xi_n | p)$:

$$\varphi(\xi_1,..., \xi_n | p) = \sum_{i=1}^{n} \xi_i .$$

Случайные величины ξ_i независимы, имеют одинаковое распределение и конечные математическое ожидание и дисперсию, поэтому для случайных величин ξ_i справедлива центральная предельная теорема, в соответствии с которой сумма $\sum_{i=1}^{n} \xi_i$ имеет асимптотически (при $n \to \infty$) нормальное распределение с параметрами $N\left(M\left[\sum_{i=1}^{n} \xi_i\right], D\left[\sum_{i=1}^{n} \xi_i\right]\right)$, где:

$$M\left[\sum_{i=1}^{n} \xi_i\right] = np ,$$

$$D\left[\sum_{i=1}^{n} \xi_i\right] = np(1 - p) ,$$

тогда случайная величина:

$$\frac{\varphi(\xi_1,\ldots,\xi_n)-np}{\sqrt{np(1-p)}}=\frac{\sum\limits_{i=1}^{n}\xi_i-np}{\sqrt{np(1-p)}}$$

имеет асимптотически (при $n\to\infty$) нормальное распределение $N(0,1)$:

$$\frac{\varphi(\xi_1,\ldots,\xi_n)-np}{\sqrt{np(1-p)}}\sim N(0,1)\text{, при } n\to\infty\ .$$

Пусть y_2 – квантиль распределения $N(0,1)$ уровня $\frac{1+P_д}{2}$, тогда при больших n:

$$P\left\{-y_2<\frac{\varphi(\xi_1,\ldots,\xi_n)-np}{\sqrt{np(1-p)}}<y_2\right\}\approx P_д\,,$$

$$P\left\{-y_2<\frac{\sum\limits_{i=1}^{n}\xi_i-np}{\sqrt{np(1-p)}}<y_2\right\}\approx P_д\,,$$

$$P\left\{-y_2<\frac{\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i-p}{\sqrt{\frac{p(1-p)}{n}}}<y_2\right\}\approx P_д\,,$$

$$P\left\{\left|\frac{\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i-p}{\sqrt{\frac{p(1-p)}{n}}}\right|<y_2\right\}\approx P_д\,,$$

$$P\left\{\left|\frac{\hat{m}_1-p}{\sqrt{\frac{p(1-p)}{n}}}\right|<y_2\right\}\approx P_д\,,$$

где $\hat{m}_1=\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i$.

Далее можно применить несколько различных способов построения приближенного доверительного интервала различных по трудоемкости и ширине получаемого интервала.

Способ А (наиболее трудоемкий): разрешая неравенство относительно неизвестной вероятности p, получим:

$$\left|\frac{\hat{m}_1-p}{\sqrt{\frac{p(1-p)}{n}}}\right|<y_2\Leftrightarrow\left(\frac{\hat{m}_1-p}{\sqrt{\frac{p(1-p)}{n}}}\right)^2<y_2^2\Leftrightarrow\frac{(\hat{m}_1-p)^2}{\frac{p(1-p)}{n}}-y_2^2<0\Leftrightarrow\frac{(\hat{m}_1-p)^2-y_2^2\frac{p(1-p)}{n}}{\frac{p(1-p)}{n}}<0\Leftrightarrow$$

$$\frac{\hat{m}_1^2-2\hat{m}_1p+p^2-\frac{y_2^2}{n}p+\frac{y_2^2}{n}p^2}{\frac{p(1-p)}{n}}<0\Leftrightarrow\frac{\left(1+\frac{y_2^2}{n}\right)p^2-\left(2\hat{m}_1+\frac{y_2^2}{n}\right)p+\hat{m}_1^2}{\frac{p(1-p)}{n}}<0\Leftrightarrow$$

$$\begin{cases} \dfrac{\left(2\hat{m}_1 + \dfrac{y_2^2}{n}\right) - \sqrt{\left(2\hat{m}_1 + \dfrac{y_2^2}{n}\right)^2 - 4\left(1 + \dfrac{y_2^2}{n}\right)\hat{m}_1^2}}{2\left(1 + \dfrac{y_2^2}{n}\right)} < p < \dfrac{\left(2\hat{m}_1 + \dfrac{y_2^2}{n}\right) + \sqrt{\left(2\hat{m}_1 + \dfrac{y_2^2}{n}\right)^2 - 4\left(1 + \dfrac{y_2^2}{n}\right)\hat{m}_1^2}}{2\left(1 + \dfrac{y_2^2}{n}\right)} \\ \dfrac{p(1-p)}{n} \neq 0 \end{cases} \Leftrightarrow$$

$$\begin{cases} \dfrac{\left(2\hat{m}_1 + \dfrac{y_2^2}{n}\right) - \sqrt{4\hat{m}_1 \dfrac{y_2^2}{n} + \dfrac{y_2^4}{n^2} - 4\dfrac{y_2^2}{n}\hat{m}_1^2}}{2\left(1 + \dfrac{y_2^2}{n}\right)} < p < \dfrac{\left(2\hat{m}_1 + \dfrac{y_2^2}{n}\right) + \sqrt{4\hat{m}_1 \dfrac{y_2^2}{n} + \dfrac{y_2^4}{n^2} - 4\dfrac{y_2^2}{n}\hat{m}_1^2}}{2\left(1 + \dfrac{y_2^2}{n}\right)} \\ \dfrac{p(1-p)}{n} \neq 0 \end{cases} \Leftrightarrow$$

$$\begin{cases} \dfrac{\left(2\hat{m}_1 + \dfrac{y_2^2}{n}\right) - \sqrt{4\dfrac{y_2^2}{n}\hat{m}_1(1-\hat{m}_1) + \dfrac{y_2^4}{n^2}}}{2\left(1 + \dfrac{y_2^2}{n}\right)} < p < \dfrac{\left(2\hat{m}_1 + \dfrac{y_2^2}{n}\right) + \sqrt{4\dfrac{y_2^2}{n}\hat{m}_1(1-\hat{m}_1) + \dfrac{y_2^4}{n^2}}}{2\left(1 + \dfrac{y_2^2}{n}\right)} \\ \dfrac{p(1-p)}{n} \neq 0 \end{cases}$$

Отсюда, приближенный доверительный интервал:

$$\left(\dfrac{\left(2\hat{m}_1 + \dfrac{y_2^2}{n}\right) - \sqrt{4\dfrac{y_2^2}{n}\hat{m}_1(1-\hat{m}_1) + \dfrac{y_2^4}{n^2}}}{2\left(1 + \dfrac{y_2^2}{n}\right)} ; \dfrac{\left(2\hat{m}_1 + \dfrac{y_2^2}{n}\right) + \sqrt{4\dfrac{y_2^2}{n}\hat{m}_1(1-\hat{m}_1) + \dfrac{y_2^4}{n^2}}}{2\left(1 + \dfrac{y_2^2}{n}\right)} \right).$$

Если в полученной системе пренебречь слагаемыми с множителем $\dfrac{1}{n}$ и с множителем $\dfrac{1}{n^2}$ под корнем, получим приближенное неравенство:

$$\begin{cases} \dfrac{2\hat{m}_1 - \sqrt{4y_2^2 \dfrac{\hat{m}_1(1-\hat{m}_1)}{n}}}{2} < p < \dfrac{2\hat{m}_1 + \sqrt{4y_2^2 \dfrac{\hat{m}_1(1-\hat{m}_1)}{n}}}{2}, \\ \dfrac{p(1-p)}{n} \neq 0 \end{cases}$$

$$\begin{cases} \dfrac{2\hat{m}_1 - 2y_2\sqrt{\dfrac{\hat{m}_1(1-\hat{m}_1)}{n}}}{2} < p < \dfrac{2\hat{m}_1 + 2y_2\sqrt{\dfrac{\hat{m}_1(1-\hat{m}_1)}{n}}}{2}, \\ \dfrac{p(1-p)}{n} \neq 0 \end{cases}$$

$$\begin{cases} \hat{m}_1 - y_2\sqrt{\dfrac{\hat{m}_1(1-\hat{m}_1)}{n}} < p < \hat{m}_1 + y_2\sqrt{\dfrac{\hat{m}_1(1-\hat{m}_1)}{n}} \\ \dfrac{p(1-p)}{n} \neq 0 \end{cases}.$$

Таким образом,

$$P\left\{ \hat{m}_1 - y_2\sqrt{\dfrac{\hat{m}_1(1-\hat{m}_1)}{n}} < p < \hat{m}_1 + y_2\sqrt{\dfrac{\hat{m}_1(1-\hat{m}_1)}{n}} \right\} \approx P_д,$$

и «приближенный» доверительный интервал для вероятности p имеет вид:

$$\left(\hat{m}_1 - y_2\sqrt{\frac{\hat{m}_1(1-\hat{m}_1)}{n}};\hat{m}_1 + y_2\sqrt{\frac{\hat{m}_1(1-\hat{m}_1)}{n}}\right).$$

Способ Б: последний интервал может быть получен несколько иным путем. Необходимость в решении громоздкого квадратного уравнения возникает из-за того, что неизвестный параметр p содержится и в числителе и в знаменателе. Замена параметра p в знаменателе некоторой оценкой, например, выборочным средним $\hat{m}_1$ избавляет от необходимости решения квадратного уравнения:

$$P\left\{\left|\frac{\hat{m}_1 - p}{\sqrt{\frac{\hat{m}_1(1-\hat{m}_1)}{n}}}\right| < y_2\right\} \approx P_д,$$

$$P\left\{-y_2 < \frac{\hat{m}_1 - p}{\sqrt{\frac{\hat{m}_1(1-\hat{m}_1)}{n}}} < y_2\right\} \approx P_д,$$

$$P\left\{\hat{m}_1 - y_2\sqrt{\frac{\hat{m}_1(1-\hat{m}_1)}{n}} < p < \hat{m}_1 + y_2\sqrt{\frac{\hat{m}_1(1-\hat{m}_1)}{n}}\right\} \approx P_д.$$

Отсюда приближенный доверительный интервал:

$$\left(\hat{m}_1 - y_2\sqrt{\frac{\hat{m}_1(1-\hat{m}_1)}{n}};\hat{m}_1 + y_2\sqrt{\frac{\hat{m}_1(1-\hat{m}_1)}{n}}\right).$$

Способ В (наименее трудоемкий): другой способ устранить зависимость знаменателя от параметра p заключается в том, чтобы весь знаменатель заменить на его оценку сверху, например, на наибольшее значение знаменателя.

Поскольку из выполнения условия $\left|\frac{\hat{m}_1 - p}{\sqrt{\frac{p(1-p)}{n}}}\right| < y_2$ следует выполнение условия $\left|\frac{\hat{m}_1 - p}{\max\limits_p \sqrt{\frac{p(1-p)}{n}}}\right| < y_2$, то имеет место вложенность событий:

$$\left\{\omega : \left|\frac{\hat{m}_1(\xi_1(\omega),\ldots,\xi_n(\omega)) - p}{\sqrt{\frac{p(1-p)}{n}}}\right| < y_2\right\} \subseteq \left\{\omega : \left|\frac{\hat{m}_1(\xi_1(\omega),\ldots,\xi_n(\omega)) - p}{\max\limits_p \sqrt{\frac{p(1-p)}{n}}}\right| < y_2\right\}.$$

Откуда следует неравенство для вероятностей:

$$P_д \approx \left\{\omega : \left|\frac{\hat{m}_1(\xi_1(\omega),\ldots,\xi_n(\omega)) - p}{\sqrt{\frac{p(1-p)}{n}}}\right| < y_2\right\} \le \left\{\omega : \left|\frac{\hat{m}_1(\xi_1(\omega),\ldots,\xi_n(\omega)) - p}{\max\limits_p \sqrt{\frac{p(1-p)}{n}}}\right| < y_2\right\}.$$

Таким образом,

$$P\left\{\left|\frac{\hat{m}_1 - p}{\max\limits_p \sqrt{\frac{p(1-p)}{n}}}\right| < y_2\right\} \approx P_\partial$$

Легко видеть, что наибольшее значение выражения $p(1-p)$ достигается при $p = \frac{1}{2}$, тогда:

$$\max_p \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{\max\limits_p p(1-p)}{n}} = \sqrt{\frac{\frac{1}{2}\left(1-\frac{1}{2}\right)}{n}} = \sqrt{\frac{1}{4n}}.$$

Таким образом,

$$P\left\{\left|\frac{\hat{m}_1 - p}{\sqrt{\frac{1}{4n}}}\right| < y_2\right\} \approx P_\partial,$$

$$P\left\{-y_2 < \frac{\hat{m}_1 - p}{\sqrt{\frac{1}{4n}}} < y_2\right\} \approx P_\partial,$$

$$P\left\{\hat{m}_1 - y_2\sqrt{\frac{1}{4n}} < p < \hat{m}_1 + y_2\sqrt{\frac{1}{4n}}\right\} \approx P_\partial.$$

Отсюда приближенный доверительный интервал имеет вид:

$$\left(\hat{m}_1 - y_2\sqrt{\frac{1}{4n}}; \hat{m}_1 + y_2\sqrt{\frac{1}{4n}}\right).$$

Нетрудно заметить, что полученный приближенный доверительный интервал является самым широким из всех полученных ранее.

8. Доверительный интервал для коэффициента корреляции двумерного нормального распределения с неизвестными математическими ожиданиями и дисперсиями.

Пусть $\left(\begin{pmatrix}\xi_1\\ \eta_1\end{pmatrix},...,\begin{pmatrix}\xi_n\\ \eta_n\end{pmatrix}\right)$ выборка из двумерного нормального распределения $N\left(\begin{pmatrix}m_\xi\\ m_\eta\end{pmatrix},\begin{pmatrix}\sigma_\xi^2 & \rho_{\xi\eta}\sigma_\xi\sigma_\eta\\ \rho_{\xi\eta}\sigma_\xi\sigma_\eta & \sigma_\eta^2\end{pmatrix}\right)$ с неизвестными математическими ожиданиями m_ξ и m_η, и неизвестными дисперсиями σ_ξ и σ_η. Требуется построить доверительный интервал для коэффициента корреляции $\rho_{\xi\eta}$ с уровнем доверия P_∂.

Если случайная величина $\begin{pmatrix}\xi\\ \eta\end{pmatrix}$ имеет распределение $N\left(\begin{pmatrix}m_\xi\\ m_\eta\end{pmatrix},\begin{pmatrix}\sigma_\xi^2 & \rho_{\xi\eta}\sigma_\xi\sigma_\eta\\ \rho_{\xi\eta}\sigma_\xi\sigma_\eta & \sigma_\eta^2\end{pmatrix}\right)$, то коэффициент корреляции $\rho_{\xi\eta}$:

$$\rho_{\xi\eta} = \frac{\operatorname{cov}(\xi,\eta)}{\sigma_\xi \cdot \sigma_\eta} = \frac{M[(\xi - m_\xi)(\eta - m_\eta)]}{\sigma_\xi \cdot \sigma_\eta}.$$

Моментная оценка коэффициента корреляции $\hat{\rho}_{\xi\eta}$ имеет вид:

$$\hat{\rho}_{\xi\eta} = \frac{\frac{1}{n}\sum_{i=1}^{n}(\xi_i - \hat{m}_\xi)(\eta_i - \hat{m}_\eta)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\xi_i - \hat{m}_\xi)^2} \cdot \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\eta_i - \hat{m}_\eta)^2}},$$

$$\hat{m}_\xi = \frac{1}{n}\sum_{i=1}^{n}\xi_i, \; \hat{m}_\eta = \frac{1}{n}\sum_{i=1}^{n}\eta_i .$$

Можно показать, что статистика $\hat{\rho}_{\xi\eta}$ имеет асимптотически нормальное распределение $N\left(\rho_{\xi\eta} - \frac{\rho_{\xi\eta}(1-\rho_{\xi\eta}^2)}{2n}, \frac{(1-\rho_{\xi\eta}^2)^2}{n-1}\right)$, однако, использовать непосредственно статистику $\hat{\rho}_{\xi\eta}$ для построения доверительного интервала весьма затруднительно, поскольку зависимость дисперсии от $\rho_{\xi\eta}$ в конечном счете приводит к необходимости решать громоздкое квадратное уравнение, и к тому же указанное асимптотическое распределение является не слишком точным для малых n и значений $\rho_{\xi\eta}$ близких к 1 и -1.

Во избежание указанных проблем прибегают к преобразованию Фишера:

$$z(\hat{\rho}_{\xi\eta}) = arth\, \hat{\rho}_{\xi\eta} = \frac{1}{2}\ln\frac{1+\hat{\rho}_{\xi\eta}}{1-\hat{\rho}_{\xi\eta}} .$$

Можно показать, что статистика z имеет асимптотически нормальное распределение $N(m_z(\rho_{\xi\eta}), \sigma_z^2(n))$, где:

$$m_z(\rho_{\xi\eta}) = \frac{1}{2}\ln\frac{1+\rho_{\xi\eta}}{1-\rho_{\xi\eta}} + \frac{\rho_{\xi\eta}}{2(n-1)} = arth\, \rho_{\xi\eta} + \frac{\rho_{\xi\eta}}{2(n-1)},$$

$$\sigma_z^2(n) = \frac{1}{n-3},$$

причем $\sigma_z^2(n)$ не зависит от $\rho_{\xi\eta}$. Легко видеть, что случайная величина $\varphi((\xi_1,\eta_1),\ldots,(\xi_n,\eta_n) \mid \rho_{\xi\eta})$:

$$\varphi((\xi_1,\eta_1),\ldots,(\xi_n,\eta_n) \mid \rho_{\xi\eta}) = \frac{z(\hat{\rho}_{\xi\eta}) - m_z(\rho_{\xi\eta})}{\sigma_z(n)} = \sqrt{n-3}(z(\hat{\rho}_{\xi\eta}) - m_z(\rho_{\xi\eta}))$$

будет иметь асимптотически нормальное распределение $N(0,1)$ при возрастании n и поэтому может быть использована для построения «приближенного» доверительного интервала. Для этого достаточно вычислить y_2 – квантиль распределения $N(0,1)$ уровня $\frac{1+P_д}{2}$, тогда:

$$P\{-y_2 < \sqrt{n-3}(z(\hat{\rho}_{\xi\eta}) - m_z(\rho_{\xi\eta})) < y_2\} \approx P_д .$$

Далее, необходимо преобразовать двойное неравенство, стоящее под знаком вероятности:

$$-y_2 < \sqrt{n-3}(z(\hat{\rho}_{\xi\eta}) - m_z(\rho_{\xi\eta})) < y_2,$$

$$-\frac{y_2}{\sqrt{n-3}} < z(\hat{\rho}_{\xi\eta}) - m_z(\rho_{\xi\eta}) < \frac{y_2}{\sqrt{n-3}},$$

$$z(\hat{\rho}_{\xi\eta}) - \frac{y_2}{\sqrt{n-3}} < m_z(\rho_{\xi\eta}) < z(\hat{\rho}_{\xi\eta}) + \frac{y_2}{\sqrt{n-3}},$$

$$z(\hat{\rho}_{\xi\eta}) - \frac{y_2}{\sqrt{n-3}} < arth\, \rho_{\xi\eta} + \frac{\rho_{\xi\eta}}{2(n-1)} < z(\hat{\rho}_{\xi\eta}) + \frac{y_2}{\sqrt{n-3}} .$$

В последнем двойном неравенстве величина $\dfrac{\rho_{\xi\eta}}{2(n-1)}$ является малой (поскольку, $|\rho_{\xi\eta}| \le 1$) и её заменяют величиной $\dfrac{\hat{\rho}_{\xi\eta}}{2(n-1)}$:

$$z(\hat{\rho}_{\xi\eta}) - \frac{y_2}{\sqrt{n-3}} < arth\,\rho_{\xi\eta} + \frac{\hat{\rho}_{\xi\eta}}{2(n-1)} < z(\hat{\rho}_{\xi\eta}) + \frac{y_2}{\sqrt{n-3}},$$

$$z(\hat{\rho}_{\xi\eta}) - \frac{y_2}{\sqrt{n-3}} - \frac{\hat{\rho}_{\xi\eta}}{2(n-1)} < arth\,\rho_{\xi\eta} < z(\hat{\rho}_{\xi\eta}) + \frac{y_2}{\sqrt{n-3}} - \frac{\hat{\rho}_{\xi\eta}}{2(n-1)}.$$

Обозначим левую и правую части неравенства статистиками $z_1(\hat{\rho}_{\xi\eta})$ и $z_2(\hat{\rho}_{\xi\eta})$:

$$z_1(\hat{\rho}_{\xi\eta}) = z(\hat{\rho}_{\xi\eta}) - \frac{y_2}{\sqrt{n-3}} - \frac{\hat{\rho}_{\xi\eta}}{2(n-1)},$$

$$z_2(\hat{\rho}_{\xi\eta}) = z(\hat{\rho}_{\xi\eta}) + \frac{y_2}{\sqrt{n-3}} - \frac{\hat{\rho}_{\xi\eta}}{2(n-1)},$$

тогда:

$$z_1(\hat{\rho}_{\xi\eta}) < arth\,\rho_{\xi\eta} < z_2(\hat{\rho}_{\xi\eta}),$$

$$thz_1(\hat{\rho}_{\xi\eta}) < \rho_{\xi\eta} < thz_2(\hat{\rho}_{\xi\eta}).$$

Таким образом,

$$P\{thz_1(\hat{\rho}_{\xi\eta}) < \rho_{\xi\eta} < thz_2(\hat{\rho}_{\xi\eta})\} \approx P_д,$$

и интервал $(th\,z_1(\hat{\rho}_{\xi\eta}), th\,z_2(\hat{\rho}_{\xi\eta}))$ является приближенным доверительным интервалом для коэффициента корреляции $\rho_{\xi\eta}$ с уровнем доверия $P_д$.

Значения функций $th\,z_1(r)$ и $th\,z_2(r)$ для различных значений $r \in [-1,1]$ сведены в таблицу, поэтому на практике после вычисления значения статистики $\hat{\rho}_{\xi\eta}$, границы доверительного интервала $th\,z_1(\hat{\rho}_{\xi\eta})$ и $th\,z_2(\hat{\rho}_{\xi\eta})$ могут быть определены из таблицы.

Рисунок 4.2. Номограмма.