Федеральное агентство железнодорожного транспорта
Государственное образовательное учреждение
Высшего профессионального образования
«ПЕТЕРБУРГСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПУТЕЙ СООБЩЕНИЯ»

Кафедра «Математика и моделирование»

М. М. ЛУЦЕНКО

ТОЧЕЧНЫЕ И ИНТЕРВАЛЬНЫЕ ОЦЕНКИ ПАРАМЕТРОВ. ПРОВЕРКА ГИПОТЕЗЫ О ВИДЕ РАСПРЕДЕЛЕНИЯ.

Методические указания к лабораторной работе
с дополнительными сведениями по теории вероятностей и математической статистике

САНКТ-ПЕТЕРБУРГ 2009

СОДЕРЖАНИЕ

Ведение

Предлагаемые методические указания должны помочь студентам при выполнении лабораторной работы и анализе полученных результатов. Лабораторная работа написана с использованием электронной таблицы MS Excel, содержит материал, традиционно включаемый в три лабораторные работы: «Первичная обработка данных», «Построение точечных и интервальных оценок для математического ожидания и дисперсии», «Проверка гипотезы о виде распределения по критерию Пирсона».

Студенты выполняют лабораторную работу в компьютерном классе университета за 2 часа. После ее выполнения должен быть составлен отчет, защищаемый на дополнительном занятии (см. пример отчета). В методических указаниях приведены вопросы для самоконтроля и необходимые сведения из теории вероятностей и статистики.

Допустимы два способа ввода исходных данных.

- По номеру варианта. В этом случае данные генерируются программой, но остается возможность сократить объем рассматриваемой выборки, уменьшая число n.
- Данные вводятся непосредственно в таблицу вариант 0 (не более 160). Эту опцию используют или при анализе малых выборок, или при выполнении других инженерных расчетов.

1. Задачи

1.1 Экономическая задача

При разработке технического задания на модернизацию асфальтобетонного завода были собраны заявки на асфальтобетон, поступившие на завод в течение n суток. Объемы суточных заявок сведены в таблицу.

- Считая, что объем суточной заявки X является случайной величиной, оценить ее среднее, дисперсию и другие числовые характеристики.
- Проверить гипотезу о нормальности распределения случайной величины (генеральной совокупности) X.
- Дать рекомендации о номинальной суточной мощности a нового предприятия и оценить возможные отклонения его суточной нагрузки.

1.2 Математическая задача

Выборка объема n из генеральной совокупности X задана в форме таблицы (по данным экономической задачи).

1. Провести сортировку и группировку выборки.

2. Найти выборочные начальные и центральные моменты до четвертого порядка включительно.

3. Оценить математическое ожидание $\mathbf{M}X = a$, дисперсию $\mathbf{D}X = \sigma^2$, среднеквадратичное отклонение σ генеральной совокупности X, используя выборочные начальные и центральные моменты.

4. Построить доверительные интервалы для математического ожидания a и дисперсии σ^2 при различных значений доверительной вероятности $\gamma = 0,9; 0,95; 0,99$.

5. Проверить гипотезу о нормальности распределения генеральной совокупности X.

- по выборочным асимметрии и эксцессу
- по соответствию гистограммы частот графику плотности нормального распределения
- по критерию Пирсона (хи-квадрат) с уровнем значимости $\alpha = 0,05$.

2. Решение математических задач

2.1 Сортировка и группировка

Пусть $x_1, x_2, \ldots, x_n$ – выборка из генеральной совокупности X. Выполним сортировку и группировку этой выборки

- Обозначим через A и B – наименьший и наибольший элементы выборки. Запишем элементы выборки в порядке возрастания:

$$A = x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)} = B$$

- Разобьем отрезок $[A,B]$ на k равных частей точками $t_0=A$, t_1, … , $t_k=B$. Шаг разбиения: $h = (B – A)/k$. Обозначим через $\overline{x_i} = (t_{i-1} + t_i)/2$ середину i-го отрезка, а через n_i – эмпирическую частоту, т.е. число точек выборки попавших в промежуток $[t_{i-1}, t_i)$ (для $i = n$ берем промежуток $[t_{n-1}, t_n]$)

2.2 Выборочные моменты

Таблица, в первой строчке которой середины промежутков группирования, а во второй относительные частоты n_i/n, может рассматриваться как таблица распределения дискретной случайной величины.

В пределах допустимой на практике точности выборочные начальные и центральные моменты порядка r для *группированной выборки* находятся по следующим формулам

$$\overline{m}_r = \frac{1}{n} \cdot \sum_{i=1}^{k} n_i \cdot \overline{x}_i^r \qquad \overline{\mu}_r = \frac{1}{n} \sum_{i=1}^{k} n_i \cdot (\overline{x}_i - \overline{x})^r ,$$

в которых выборочное среднее равно

$$\overline{m}_1 = \frac{1}{n} \cdot \sum_{i=1}^{k} n_i \cdot \overline{x}_i = \overline{x} .$$

Напомним, что для не группированной выборки выборочное среднее находится как среднее арифметическое всех элементов выборки:

$$\overline{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

Первые четыре центральных момента μ_r случайной величины могут быть выражены через начальные моменты m_r той же случайной величины по формулам:

$$\mu_1 = 0;$$
$$\mu_2 = m_2 - m_1^2\ ;$$
$$\mu_3 = m_3 - 3m_2 \cdot m_1 + 2m_1^3;$$
$$\mu_4 = m_4 - 4m_3 \cdot m_1 + 6m_2 \cdot m_1^2 - 3m_1^4.$$

Последние формулы можно использовать и при вычислении выборочных центральных моментов $\overline{\mu}_r$ по выборочным начальным моментам $\overline{m}_r$.

2.3 Асимметрия и эксцесс генеральной совокупности

Центральный момент третьего порядка случайной величины характеризует несимметричность распределения относительно математического ожидания. Поэтому за характеристику несимметричности распределения принимают безразмерную величину – отношение третьего центрального момента к кубу среднеквадратичного отклонения

$$g_1 = \mu_3 / \sigma^3 = \mu_3 / \mu_2^{3/2}\ .$$

Эта величина называется *асимметрией* или скошенностью распределения случайной величины.

Центральный момент четвертого порядка при данной дисперсии может служить характеристикой удельного веса больших отклонений от математического ожидания, а это в свою очередь определяет характер максимума в точке *a* симметричного распределения – "островершинность" или "плосковершинность" плотности распределения. Поэтому за характеристику поведения кривой распределения вблизи точки *a* принимают безразмерную величину

$$g_2 = \mu_4 / \mu_2^2 - 3.$$

Эта величина называется *эксцессом* распределения случайной величины.

Напомним, что у нормально распределенной случайной величины асимметрия и эксцесс равны нулю.

2.4 Точечные оценки математического ожидания и дисперсии

Состоятельной и практически несмещенной оценкой математического ожидания **M**X=a генеральной совокупности X будет выборочное среднее:

$$\overline{x} = \frac{1}{n} \cdot \sum_{i=1}^{k} n_i \cdot \overline{x}_i\ ,$$

то есть, верны следующие равенства

$$\lim_{n\to\infty} \overline{x} = a; \qquad \mathrm{M}\overline{x} \approx a\quad .$$

Состоятельной и практически несмещенной оценкой дисперсии $\mathbf{D}X=\sigma^2$ генеральной совокупности X будет выборочная дисперсия:

$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{k} n_i \cdot \left(\bar{x}_i - \bar{x}\right)^2 ,$$

то есть, верны следующие равенства

$$\lim_{n\to\infty} s^2 = \sigma^2 ; \qquad \mathrm{M}s^2 \approx \sigma^2 \quad .$$

Квадратный корень из выборочной дисперсии, то есть s, называют выборочным среднеквадратичным отклонением.

2.5 Доверительные интервалы для математического ожидания

Границы доверительного интервала $[a_1,a_2]$ для математического ожидания a генеральной совокупности X находятся по следующим формулам

$$a_1 = \bar{x} - t_\gamma \frac{s}{\sqrt{n}} \qquad a_2 = \bar{x} + t_\gamma \frac{s}{\sqrt{n}} ,$$

в которых $\bar{x}$ – выборочное среднее, s – выборочное среднеквадратичное отклонение, n – объем выборки, γ – доверительная вероятность. При большом объеме выборки число t_γ находят из решения следующего уравнения

$$\Phi_0(t_\gamma) = \gamma / 2 ,$$

где, $\Phi_0(t)$ – интегральная функция Лапласа:

$$\Phi_0(t) = \frac{1}{\sqrt{2\pi}} \cdot \int_0^t \mathrm{e}^{-z^2/2} dz$$

Таким образом, оцениваемое значение математического ожидания a, окажется внутри отрезка $[a_1,a_2]$ с вероятностью γ.

Если объем выборки небольшой, то t_γ находится из решения уравнения $F(t_\gamma) = (1-\gamma)/2$, где γ как и ранее, доверительная вероятность, а $F(x)$ – функция распределения Стьюдента величины $\tau(n\text{-}1)$ с $(n-1)$-й степенью свободы (см. табл. Стьюд.).

2.6 Доверительные интервалы для дисперсии.

Построим доверительный интервал для дисперсии. Предполагая, что генеральная совокупность X имеет нормальное распределение с параметрами a и σ, можно доказать, что статистика

$$\frac{1}{\sigma^2} \sum_{i=1}^{k} n_i \cdot \left(\bar{x}_i - \bar{x}\right)^2$$

имеет распределение близкое к $\chi^2(l)$ с $l = n-1$ степенями свободы. Следовательно,

$$\mathrm{P}\left(\beta_1 < \frac{n-1}{\sigma^2}\cdot s^2 < \beta_2\right) = \mathrm{P}\left(\beta_1 < \chi^2(l) < \beta_2\right) = \frac{1}{\Gamma(l/2)2^{l/2}}\cdot\int_{\beta_1}^{\beta_2} x^{\frac{l-1}{2}}\cdot \mathrm{e}^{-x/2}\,dx = \gamma$$

где γ, как и ранее, доверительная вероятность. Числа β_1 и β_2 будут решениями уравнения

$$\mathrm{P}\left(\chi^2(l) < \beta\right) = \frac{1}{\Gamma(l/2)2^{l/2}}\cdot\int_0^{\beta} x^{\frac{l-2}{2}}\cdot \mathrm{e}^{-x/2}\,dx = \varepsilon$$

для $\varepsilon=(1-\gamma)/2$ и $\varepsilon=(1+\gamma)/2$ соответственно. Для решения этих уравнений можно воспользоваться таблицами квантилей распределения χ^2 (см. Табл. ХИ). Границы доверительного интервала для σ^2 будут равны

$$b_1 = \frac{s^2}{\beta_2}\cdot(n-1) \qquad b_2 = \frac{s^2}{\beta_1}\cdot(n-1)$$

Итак, оцениваемая дисперсия σ^2 с вероятностью γ лежит в промежутке $[b_1, b_2]$.

2.7 Проверка гипотезы о виде распределения

2.7.1 Формулировка гипотезы.

Предположим, что генеральная совокупность X имеет нормальное распределение $N(\bar{x}, s)$ с математическим ожиданием $\bar{x}$ и среднеквадратичным отклонением s.

Проверим это предположение по следующим трем критериям.

2.7.2 Проверка визуального соответствия

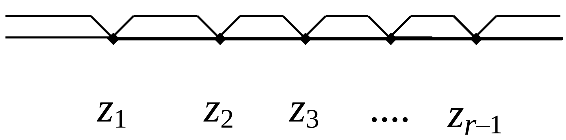
Гистограмма выборочных частот и кривая плотности нормального распределения строятся в одинаковом масштабе. Если эти графики "сильно" отличаются друг от друга, то гипотеза отвергается.

2.7.3 Сравнение выборочных моментов с теоретическими

По выборочным центральным моментам найдем: выборочные асимметрию и эксцесс. Если найденные значения "значительно" отличаются от нуля, то наша гипотеза неверна.

2.7.4 Критерий согласия χ^2

Данный критерий позволяет количественно оценить степень согласия выборочных данных проверяемой гипотезе.



Выполним новую группировку первоначальной выборки.

Для этого всю числовую ось разобьем на r возможно неравных частей, так, что внутри каждой части окажется не менее 5 элементов выборки. Обозначим через $-\infty < z_1 < z_2 < ... < z_{r-1} < \infty$ новые точки разбиения числовой оси. При этом мы будем рассматривать и два неограниченных промежутка: $(-\infty, z_1)$, $[z_{r-1}, \infty)$. Если рассматриваемая выборка была уже группирована, то следует объединить соседние промежутки так, чтобы в объединенные промежутки попало не менее 5 элементов выборки, а границы первого и последнего промежутка заменить на $-\infty$ и ∞.

Исходя из нашего предположения о нормальности генеральной совокупности, вычислим вероятность попадания случайной величины X в промежуток $[z_{j-1}, z_j)$.

$$p_j = \mathrm{P}\left(z_{j-1} \le \mathbf{X} < z_j\right) = \Phi_0\left(\frac{z_j - \bar{x}}{s}\right) - \Phi_0\left(\frac{z_{j-1} - \bar{x}}{s}\right)$$

Сравним вычисленную вероятность p_j с относительной выборочной частотой n_j/n, где через n_j обозначено число точек выборки, попавших в промежуток $[z_{j-1}, z_j)$. Если наша гипотеза верна, то величина $(n_j/n - p_j)^2$ должна быть небольшой. А следовательно, небольшой должна быть и сумма r таких величин взятых с коэффициентами n/p_j :

$$Z = \sum_{j=1}^{r} \frac{n}{p_j}\left(p_j - \frac{n_j}{n}\right)^2 = \sum_{j=1}^{r} \frac{\left(n_j - n \cdot p_j\right)^2}{n \cdot p_j}.$$

При больших n статистика Z имеет распределение близкое к распределению $\chi^2(l)$ с $l = r-3$ степенями свободы.

Итак, если гипотеза верна, то вероятность, с которой статистика Z будет меньше z можно найти по формуле

$$\mathrm{P}(\mathbf{Z} < z) = \mathrm{P}\left(\chi^2(l) < z\right) = \frac{1}{\Gamma\left(l/2\right)2^{l/2}} \cdot \int_0^z x^{\frac{l-2}{2}} \cdot \mathrm{e}^{-x/2}\, dx = \gamma$$

Если последнее уравнение решить относительно z, то его решение $z(\gamma)=\chi^2_{\text{крит}}$ называется квантилью распределения $\chi^2(l)$ с l степенями свободы и считают критическим значением выборочной статистики Z, соответствующим уровню значимости $\alpha=1-\gamma$. Значения квантилей распределения хи-квадрат для различного числа степеней свободы можно найти в таблицах.

Окончательно, если после проведения дополнительного группирования выборки значение статистики Z будет больше $\chi^2_{\text{крит}}$, то гипотеза отвергается. Причем вероятность отвергнуть правильную гипотезу будет не более чем α. Если же значение выборочной статистики Z будет не больше $\chi^2_{\text{крит}}$, то у нас *нет основания отвергнуть* гипотезу и она, как правило, принимается.

3. Дополнительные сведения по теории вероятностей и статистике

Случайной величиной называется величина, значения которой зависят от случая, то есть функция, отображающая множество элементарных событий опыта во множество вещественных чисел.

Функцией распределения случайной величины X называется функция:

$$F(x) = \mathbf{P}(X < x).$$

Значение функции распределения в точке x равно вероятности того, что значение случайной величины X меньше числа x.

Случайная величина X называется *дискретной*, если множество ее значений состоит из изолированных точек, если множество значений заполняет некоторый промежуток, то случайная величина X называется *непрерывной*.

Если функция распределения $F(x)$ дифференцируема, то ее производная $f(x) = F'(x)$ называется *функцией плотности распределения случайной величины* X.

Если случайная величина дискретна, то *таблицей распределения случайной величины* X называется таблица, в первой строчке которой перечислены значения случайной величины, а во второй вероятности, с которыми эти значения принимаются.

Начальным моментом случайной величины X порядка r называется следующее число:

- $\mathbf{M}X^r = \int\limits_{-\infty}^{\infty} x^r f(x)dx$, где $f(x)$ плотность распределения непрерывной случайной величины.
- $\mathbf{M}X^r = \sum\limits_{i=1}^{m} x_i^r p_i$, где x_i $i = \overline{1,m}$ – все возможные значения дискретной случайной величины, а p_i $i = \overline{1,m}$ – вероятности, с которыми эти значения принимаются.

Математическим ожиданием случайной величины X называется начальный момент этой величины первого порядка, то есть $\mathbf{M}X$.

Центральным моментом случайной величины X порядка r называется:

$$\mu_r = \mathbf{M}(X - \mathbf{M}X)^r.$$

Центральный момент второго порядка называется *дисперсией случайной величины* X:

$$\mathbf{D}X = \mathbf{M}(X - \mathbf{M}X)^2.$$

Случайная величина X *распределена по нормальному закону* с параметрами a, σ ($N(a,\sigma)$) если ее плотность распределения имеет вид:

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\cdot e^{-(x-a)^2/2\sigma^2} \text{ или } f(x)=\frac{1}{\sigma}\cdot\varphi\left(\frac{x-a}{\sigma}\right),$$

где $\varphi(t)$ – локальная функция Лапласа.

Для нормально распределенной случайной величины: $X \sim N(a,\sigma)$ математическое ожидание и дисперсия равны: $\mathbf{M}X=a$, $\mathbf{D}X=\sigma^2$.

Последовательность случайных величин $X_1, X_2, \ldots, X_n, \ldots$ *сходится по вероятности* к числу a, $X_n \to a$ если

$$\mathbf{P}(|X_n-a|<\varepsilon)\to 1 \text{ при } n\to\infty \text{ для всех } \varepsilon>0.$$

(Для любого числа $\varepsilon>0$ вероятность того что значение случайной величины отклонится от числа меньше чем на стремится к единице, или, почти наверное, абсолютная ошибка от замены числа a значением случайной величины X_n меньше чем ε при больших значениях n.)

Пусть X – случайная величина, а $F(x)$ ее функция распределения. *Квантилью распределения* X *уровня* p называется решение уравнения $F(x_p)=p$.

Выборкой объема n *из генеральной совокупности* X *с функцией распределения* $F(x)$ называется последовательность $x_1, x_2, \ldots, x_n$ наблюдаемых значений случайной величины X, соответствующих n независимым повторениям опыта. Выборку также понимают как n независимых реализаций случайной величины X.

Группированной выборкой объема n из генеральной совокупности X называется таблица, в первой строке которой взаимно непересекающиеся промежутки, а во второй – число элементов значений случайной величины X, попавших в соответствующий промежуток (промежутки выбраны, так что их объединение Δ есть промежуток, в который попадают все значения X).

Промежутки:	Δ_1	Δ_2	Δ_3	$\ldots$	Δ_m
Частоты:	n_1	n_2	n_3	$\ldots$	n_m

Гистограммой частот группированной выборки называется график кусочно-постоянной функции, значения которой на промежутках группирования равны соответствующим частотам, а в остальных точках равны нулю.

Гистограммой относительных частот группированной выборки называется график кусочно-постоянной функции, значения которой на i-ом промежутке группирования равны соответствующей относительной частоте: n_i/n, а вне промежутков группирования равны нулю.

Произвольную функцию $\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots, x_n)$ элементов выборки $x_1, x_2, \ldots, x_n$ называют *статистикой*. Если статистика $\hat{\theta}$ используется для оценки параметра θ генеральной совокупности X, то она называется *оценкой этого параметра* или просто *оценкой*.

Оценка $\hat{\theta}$ параметра θ называется *несмещенной*, если ее математическое ожидание равно оцениваемому параметру $\mathbf{M}\hat{\theta} = \theta$.

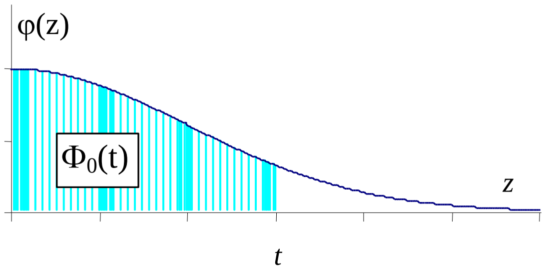
Оценка $\hat{\theta} = \hat{\theta}_n$ (последовательность оценок) параметра θ называется *состоятельной*, если при увеличении объема выборки n оценка $\hat{\theta}_n$ сходится по вероятности к оцениваемому параметру θ: $\hat{\theta}_n \to \theta$ по вероятности.

Пусть $\hat{\theta}_1$ и $\hat{\theta}_2$ – две различные несмещенные оценки параметра θ. Если $\mathbf{D}\hat{\theta}_1 < \mathbf{D}\hat{\theta}_2$, то говорят, что оценка $\hat{\theta}_1$ более *эффективна*, чем оценка $\hat{\theta}_2$

Доверительным интервалом для параметра θ называется интервал $\Delta = (\theta_1, \theta_2)$, содержащий истинное значение параметра θ с вероятностью $\gamma = 1 - \alpha$. Число γ называется *доверительной вероятностью*, а число α – *уровнем значимости*. Таким образом, $\mathbf{P}(\theta \in \Delta = (\theta_1, \theta_2)) = \gamma$.

4. Интегральная функция Лапласа

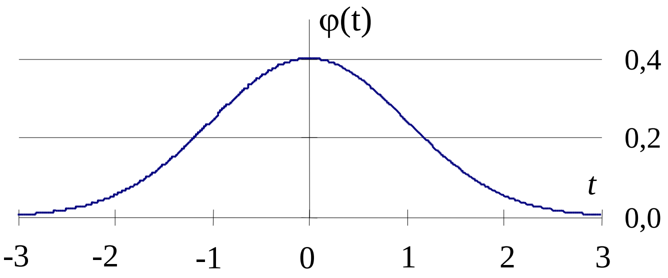
$$\Phi_0(t)=\frac{1}{\sqrt{2\pi}}\cdot\int_0^t e^{-z^2/2}dz$$



t	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
4,0	0,4999683			4,5	0,4999966			5,0	0,4999997	

5. Локальная функция Лапласа

$$\varphi(t)=\frac{1}{\sqrt{2\pi}}\cdot e^{-t^2/2}$$

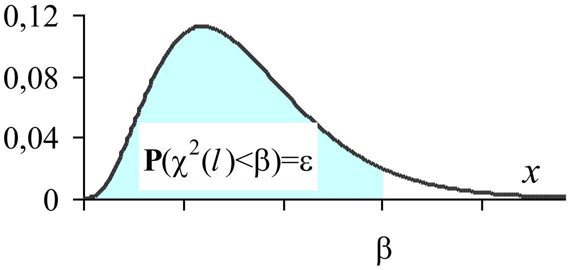


t	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3725	0,3712	0,3697
0,4	0,3683	0,3668	0,3653	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0396	0,0387	0,0379	0,0371	0,0363
2,2	0,0355	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009

3,5	0,000872683	3,6	0,000611902	3,7	0,000424780
3,8	0,000291947	3,9	0,000198655	4,0	0,000133830

6. Квантили распределения хи-квадрат

$$P\left(\chi^2(l) < \beta\right) = \frac{1}{\Gamma(l/2)\cdot 2^{l/2}} \cdot \int_0^{\beta} x^{\frac{l-2}{2}} \cdot e^{-x/2}\, dx = \varepsilon$$



	Вероятности ε					
l	**0,005**	**0,025**	**0,05**	**0,95**	**0,975**	**0,995**
1	0,00	0,00	0,00	3,84	5,02	7,88
2	0,01	0,05	0,10	5,99	7,38	10,60
3	0,07	0,22	0,35	7,81	9,35	12,84
4	0,21	0,48	0,71	9,49	11,14	14,86
5	0,41	0,83	1,15	11,07	12,83	16,75
6	0,68	1,24	1,64	12,59	14,45	18,55
7	0,99	1,69	2,17	14,07	16,01	20,28
8	1,34	2,18	2,73	15,51	17,53	21,96
9	1,73	2,70	3,33	16,92	19,02	23,59
10	2,16	3,25	3,94	18,31	20,48	25,19
11	2,60	3,82	4,57	19,68	21,92	26,76
12	3,07	4,40	5,23	21,03	23,34	28,30
13	3,57	5,01	5,89	22,36	24,74	29,82
14	4,07	5,63	6,57	23,68	26,12	31,32
15	4,60	6,26	7,26	25,00	27,49	32,80
16	5,14	6,91	7,96	26,30	28,85	34,27
17	5,70	7,56	8,67	27,59	30,19	35,72
18	6,26	8,23	9,39	28,87	31,53	37,16
19	6,84	8,91	10,12	30,14	32,85	38,58
20	7,43	9,59	10,85	31,41	34,17	40,00
21	8,03	10,28	11,59	32,67	35,48	41,40
22	8,64	10,98	12,34	33,92	36,78	42,80
24	9,89	12,40	13,85	36,42	39,36	45,56
25	10,52	13,12	14,61	37,65	40,65	46,93
49	27,25	31,55	33,93	66,34	70,22	78,23
50	27,99	32,36	34,76	67,50	71,42	79,49
99	66,51	73,36	77,05	123,23	128,42	138,99
100	67,33	74,22	77,93	124,34	129,56	140,17
149	108,29	117,10	121,79	178,49	184,69	197,21
150	109,14	117,98	122,69	179,58	185,80	198,36
199	151,37	161,83	167,36	232,91	239,96	254,14
200	152,24	162,73	168,28	233,99	241,06	255,26
249	195,28	207,19	213,47	286,81	294,60	310,23
250	196,16	208,10	214,39	287,88	295,69	311,35

7. Пример выполнения работы

7.1 Задание на лабораторную работу

1. Введите исходные данные в соответствии с номером Вашего варианта.

2. Подберите шаг группирования и выполните группировку выборки. Запишите результат группирования.

3. Постройте гистограмму частот. Сравните построенную гистограмму с графиком приведенной плотности нормального распределения.

4. Найдите выборочные начальные и центральные моменты.

5. Укажите точечные оценки начальных и центральных моментов, математического ожидания, дисперсии, асимметрии и эксцесса.

6. Вычислите доверительные интервалы для математического ожидания и дисперсии для доверительных вероятностей $\gamma = 0{,}9;\ 0{,}95;\ 0{,}99$.

7. Проверьте соответствие выборки нормальному закону по критерию хи-квадрат при уровне значимости $\alpha = 0{,}05$.

8. Отчет о лабораторной работе должен содержать:

– наилучшую группированную выборку;

– точечные оценки моментов и других характеристик;

– доверительные интервалы для математического ожидания и дисперсии;

– гистограмму частот и график приведенной плотности нормального распределения с соответствующими значениями параметров;

– выводы о согласии выборочных данных с гипотезой о нормальности распределения генеральной совокупности;

рекомендации о номинальной суточной мощности a нового предприятия и оценку возможных отклонений его суточной нагрузки.

7.2 Ввод исходных данных и сортировка

Если исходные данные Вам даны в форме таблицы, то введите их в таблицу 1. В противном случае введите номер вашего варианта на листе: "***Выполнение***".

Лист "*Выполнение"*

Введите номер Вашего варианта или 0, если исходные данные Вы уже ввели.

Вариант № *0* . В работе можно изменять только числа помеченные красным цветом (в данном тексте они помечены курсивом). Элементы выборки записаны в таблице. Из них будет обработано только первые *n= 150* элементов. Если вы хотите обрабатывать другое число элементов выборки, то измените значение *n*.

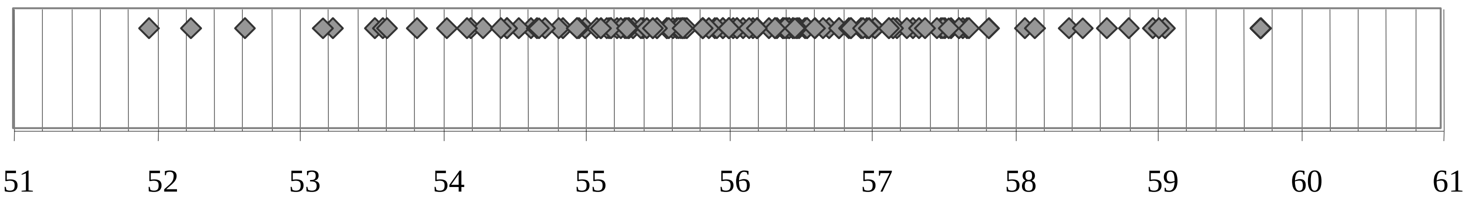
Сортировка (ранжирование)

Составим вариационный ряд, то есть запишем элементы выборки в порядке возрастания:

$$x_{(1)} < x_{(2)} < \ldots < x_{(n)}.$$

$x_{(1)} = A = 42{,}22$; $x_{(n)} = B = 49{,}82$ Выборочное среднее = 46,1077

На следующей диаграмме дано визуальное представление ранжированной выборки:



Здесь символами: ◊ обозначены элементы выборки. Для выполнения группировки перейдите на лист ***Группировка***

Вариант № О **Таблица 1**

Номер элемента	Эл-т выборки	Номер элемента	Эл-т выборки	Номер элемента	Эл-т выборки	Номер элемента	Эл-т выборки	Номер элемента	Эл-т выборки	Номер элемента	Эл-т выборки	Номер элемента	Эл-т выборки	Номер элемента	Эл-т выборки
1	*54,62*	21	*55,7*	41	*56,41*	61	*55,5*	81	*56,11*	101	*57,25*	121	*56,33*	141	*55,43*
2	*54,54*	22	*57,03*	42	*55,09*	62	*56,7*	82	*56,29*	102	*58,97*	122	*57,38*	142	*55,69*
3	*55*	23	*56,2*	43	*54,82*	63	*59,73*	83	*57,46*	103	*54,18*	123	*56,98*	143	*56,48*
4	*56,61*	24	*53,53*	44	*53,82*	64	*58,65*	84	*58,38*	104	*56,06*	124	*58,81*	144	*56,47*
5	*55,4*	25	*57,49*	45	*55,61*	65	*56,67*	85	*56,18*	105	*57,15*	125	*53,62*	145	*56,99*
6	*59,06*	26	*57,12*	46	*55,31*	66	*55,15*	86	*56,85*	106	*56,86*	126	*54,41*	146	*56,61*
7	*56,39*	27	*55,96*	47	*55,91*	67	*55,18*	87	*54,28*	107	*55,31*	127	*59,01*	147	*57,54*
8	*56,86*	28	*52,25*	48	*57,56*	68	*55,22*	88	*55,68*	108	*55,44*	128	*56,15*	148	*55,47*
9	*55,11*	29	*56,2*	49	*55,67*	69	*57,5*	89	*56,06*	109	*57,33*	129	*53,24*	149	*57,83*
10	*57,29*	30	*55,47*	50	*55,39*	70	*54,96*	90	*53,59*	110	*55,82*	130	*54,72*	150	*55,11*
11	*55,68*	31	*54,94*	51	*56,93*	71	*57,62*	91	*56,5*	111	*55,59*	131	*56,94*	151	
12	*55,34*	32	*55,83*	52	*55,97*	72	*55,25*	92	*56,34*	112	*56*	132	*57,68*	152	
13	*56,39*	33	*54,04*	53	*55,83*	73	*54,45*	93	*51,95*	113	*55,3*	133	*54,94*	153	
14	*54,2*	34	*57,52*	54	*55,65*	74	*54,45*	94	*55,92*	114	*57,03*	134	*56,48*	154	
15	*55,17*	35	*56,54*	55	*56,42*	75	*55,57*	95	*56,04*	115	*57,83*	135	*55,51*	155	
16	*54,84*	36	*58,07*	56	*55,28*	76	*57,18*	96	*56,01*	116	*55,87*	136	*55,83*	156	
17	*56,43*	37	*56,78*	57	*56,33*	77	*57,67*	97	*57,61*	117	*56,01*	137	*56,2*	157	
18	*56,38*	38	*54,67*	58	*56,46*	78	*56,55*	98	*56,97*	118	*56,55*	138	*54,68*	158	
19	*55,58*	39	*57,15*	59	*52,62*	79	*56,78*	99	*56,4*	119	*55,72*	139	*53,17*	159	
20	*57,65*	40	*56,45*	60	*58,14*	80	*55,66*	100	*55,18*	120	*57,12*	140	*58,48*	160	

7.3 Группировка

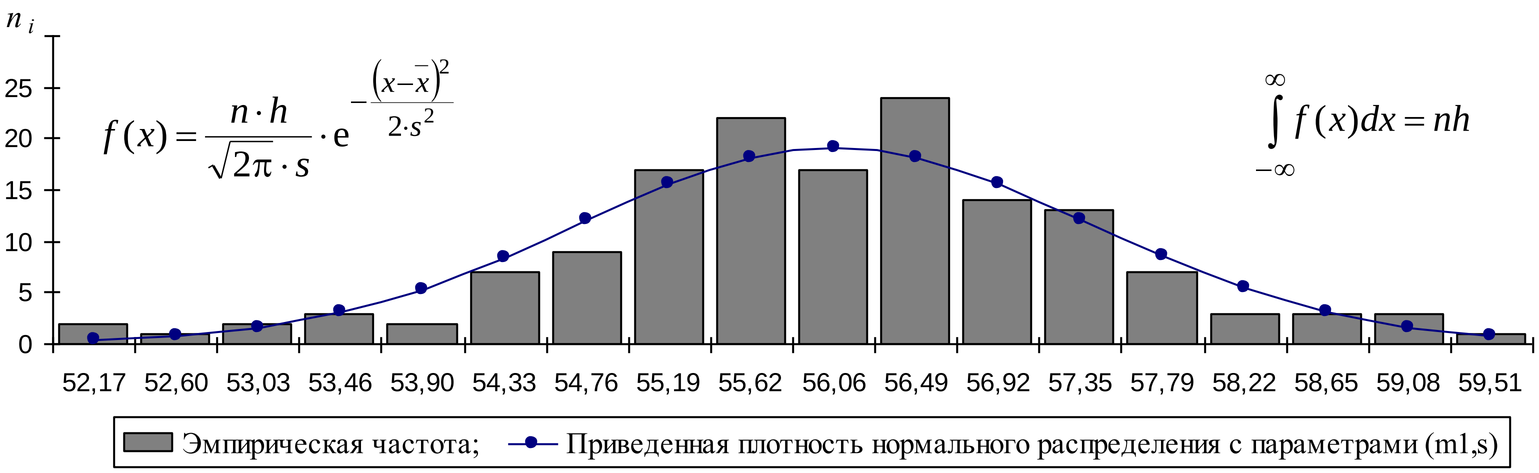
Для проведения группировки выборки выберем число k. Разобьем отрезок $[x_{(1)}, x_{(n)}]$ на k равных частей. Число элементов выборки, попавших в каждый элемент разбиения, называется частотой n_i.

$n = 150$; $k = 18$; $x_{(1)} = A = 51{,}95$; $x_{(n)} = B = 59{,}73$. Шаг разбиения $h = 0{,}4322$, размах выборки $B–A = 7{,}78$.

7.3.1 Результат группирования

Номера интервалов	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Границы интервалов	51,95	52,38	52,81	53,25	53,68	54,11	54,54	54,98	55,41	55,84	56,27	56,70	57,14	57,57	58,00	58,43	58,87	59,30
	52,38	52,81	53,25	53,68	54,11	54,54	54,98	55,41	55,84	56,27	56,70	57,14	57,57	58,00	58,43	58,87	59,30	59,73
Середины ин-лов	52,17	52,60	53,03	53,46	53,90	54,33	54,76	55,19	55,62	56,06	56,49	56,92	57,35	57,79	58,22	58,65	59,08	59,51
Частоты n_i	2	1	2	3	2	7	9	17	22	17	24	14	13	7	3	3	3	1

Гистограмма частот и график приведенной плотности нормального распределения



Для данной группированной выборки мы получим: выборочное математическое ожидание $\bar{x}$=56,06; выборочную дисперсию s^2 = 1,840; выборочное среднеквадратичное отклонение s = 1,36. На графике изображена приведенная функция плотности нормального распределения, отличающаяся от истинной функции в nh раз.

7.3.2 Выборочные начальные и центральные моменты:

m_1 = 56,06187407; m_2 = 3144,761998; m_3 = 176505,7257; m_4 = 9912404,708;

μ_1 = 0; μ_2 = 1,828273192; μ_3 = 0,518238519; μ_4 = 11,67477566.

Асимметрия: –0,207544184; Эксцесс: 0,44632619.

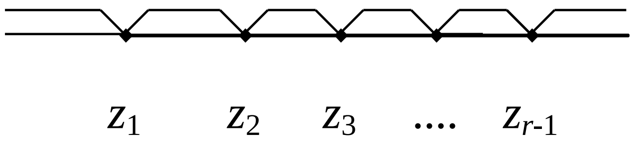
7.3.3 Доверительные интервалы для математического ожидания и дисперсии.

Квантили нормального распределения находятся по Табл. Ф0(t); квантили β_1, β_2 распределения хи-квадрат для $(n-1)$ степени свободы находятся по Табл_ХИ

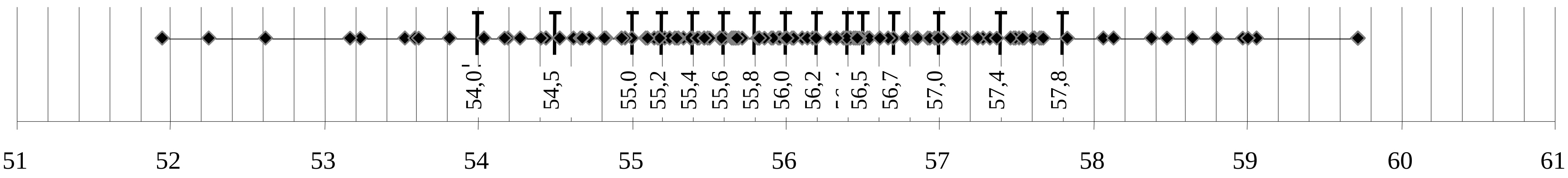
γ	t_γ	$\delta = t_\gamma \frac{s}{\sqrt{n}}$	$a_1=m_1-\delta$	$a_2=m_1+\delta$	β_1	β_2	$b_1 = \frac{s^2}{\beta_2}\cdot(n-1)$	$b_2 = \frac{s^2}{\beta_1}\cdot(n-1)$
0,9	1,645	0,18	55,88	56,24	122	178	1,536	2,252
0,95	1,95	0,22	55,85	56,28	117	185	1,485	2,342
0,99	2,575	0,29	55,78	56,35	108	197	1,391	2,532

7.4 Проверка гипотезы о нормальном распределении генеральной совокупности по критерию хи-квадрат.

Выполните новую группировку выборки. Вставьте правые границы, начиная с z_1. Причем за последнюю правую границу надо взять наибольшее значение выборки или большее его число. Помните, что $z_0 = -\infty$ и $z_r = \infty$. При новой группировке стремитесь к тому, чтобы во все интер-



валы попало примерно равное число элементов выборки, но не менее пяти. На рисунке вертикальные отметки указывают границы выбранных интервалов.



Номер ин-ла j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
z_{j-1}	$-\infty$	54,00	54,50	55,00	55,20	55,40	55,60	55,80	56,00	56,20	56,40	56,50	56,70	57,00	57,40	57,80	60,00	0,00
z_j	*54,0*	*54,5*	*55,0*	*55,2*	*55,4*	*55,6*	*55,8*	*56,0*	*56,2*	*56,4*	*56,5*	*56,7*	*57,0*	*57,4*	*57,8*	*60,0*		
частоты n_j	9	7	11	7	9	9	9	10	11	8	9	7	10	11	11	12	0	0

Вычислим r–1 значений функции $\Phi_0(u_j)$ на правых концах интервалов. Значение $\Phi_0(u_r)=\Phi_0(+\infty)$ полагаем равным 0,5.

$u_j = (z_j-m_1)/s$	$-\infty$	-1,52	-1,15	-0,78	-0,64	-0,49	-0,34	-0,19	-0,05	0,10	0,25	0,32	0,47	0,69	0,99	1,28	0,00	0,00
$\Phi_0(u_{j+1})$	*-0,436*	*-0,375*	*-0,282*	*-0,239*	*-0,188*	*-0,133*	*-0,075*	*-0,020*	*0,040*	*0,099*	*0,126*	*0,181*	*0,255*	*0,339*	*0,400*	*0,500*		
p_j	0,064	0,061	0,093	0,043	0,051	0,055	0,058	0,055	0,06	0,059	0,027	0,055	0,074	0,084	0,061	0,1	0	0
np_j	9,638	9,122	13,89	6,509	7,647	8,229	8,659	8,311	8,965	8,832	4,021	8,296	11,11	12,6	9,122	15,04	0	0

Таблица значений функции Φ_0(t) расположена на листе Табл_Фо

Разности между наблюдаемым n_j и ожидаемым значениями np_j (n_j – число точек попавших в j-й интервал, p_j – вероятность того, что значения случайной величины $N(\bar{x}, s)$ попадут в j-й интервал):

n_j-np_j	-0,64	-2,12	-2,89	0,49	1,35	0,77	0,34	1,69	2,03	-0,83	4,98	-1,30	-1,11	-1,60	1,88	-3,04	0,00	

$$Z = \sum_{j=1}^{r} \frac{\left(n_j - n \cdot p_j\right)^2}{n \cdot p_j} = 10{,}07$$

Полученное значение сравниваем с квантилю распределения $\chi^2(r–3)$ по Табл_ХИ. Соответствующее значение для доверительной вероятности $\varepsilon = 0{,}95$ равно $\chi^2(16–3) = \chi^2(13)= 22{,}36$.

7.5 Математические выводы

1. Для выборки объема n=150 найдены значения выборочных начальных и центральных моментов (см. выше).

2. Найдены точечные оценки математического ожидания $\mathbf{M}X = a$, дисперсии $\mathbf{D}X = \sigma^2$, среднеквадратичного отклонения σ генеральной совокупности X (см. выше).

3. Построены доверительные интервалы для математического ожидания a и дисперсии σ^2 для различных значений доверительной вероятности $\gamma = 0{,}9;\ 0{,}95;\ 0{,}99$ (см. таблицу)

4. Нет основания отвергнуть гипотезу о нормальности распределения генеральной совокупности X, так как

– гистограмма частот и график плотности нормального распределения "близки" друг к другу;

– выборочные асимметрия и эксцесс имеют значения "близкие" к нулю;

– по критерию Пирсона (хи-квадрат) критическое значение $\chi^2_{крит} = \chi^2_{0,95}(13) = 22{,}36 > \chi^2_{расчет} = Z = 10{,}07$

7.6 Решение задачи планирования

При разработке технического задания на строительство нового асфальтобетонного завода необходимо учесть следующее.

1. Средне дневной объем суточной заявки составит примерно $\bar{x}$=56,062 при выборочной дисперсии s^2 = 1,8405; и выборочном среднеквадратичном отклонении s = 1,357.

2. Нет основания для отвержения гипотезы о нормальности распределения генеральной совокупности X (объема суточной заявки) ни по гистограмме частот, ни по выборочным асимметрии и эксцессу, ни по критерию Пирсона с уровнем значимости $\alpha = 0{,}05$ (доверительная вероятность ($\varepsilon = 1 - \alpha = 0{,}95$).

3. На основание 150 наблюдений за поступающими заявками средний объем заявок за день будет лежать в промежутке [55,88 ; 56,24]с вероятностью 0,9; в промежутке [55,85 ; 56,28]с вероятностью 0,95, в промежутке [55,78 ; 56,35]с вероятностью 0,99. (Средний объем заявки окажется в соответствующих промежутках с соответствующими вероятностями.)

7.7 Вопросы для самоконтроля

1. Сколько раз вы проводили группировку выборки? Укажите результаты группирования.
2. Чему равны $f(\bar{x})$, $f(\bar{x} + hs)$ для приведенных функций распределения (см. п. 7.3.1)?
3. Почему выборочное среднее п. 7.2 отличается от первого начального момента п. 7.3.2.?
4. Что значат записи $\chi^2(5) = 7$, и $\chi^2_{0,95}(5) = 11{,}07$?
5. Можно ли определить значения $\mathbf{M}X$ и $\mathbf{D}X$ абсолютно точно и что для этого необходимо.
6. Каковы вероятности того, что $\mathbf{M}X = \bar{x}$ и $\mathbf{D}X = s^2$
7. Статистики $\bar{\mu}_2$ и s^2 являются оценками одной и той же характеристики случайной величины. Почему их значения различны?
8. Найти доверительные интервалы для математического ожидания и дисперсии при доверительной вероятности $\gamma = 0{,}92$.
9. Верно ли, что генеральная совокупность имеет распределение $N(\bar{x}, s)$? Какова вероятность этого события?
10. Оцените математическое ожидание для выборки 1, 3, 5, 7, 9. Укажите ее объем. Найдите доверительный интервал для математического ожидания, если известна дисперсия генеральной совокупности $\sigma^2 = 0{,}81$.