

УДК 004:519.2
МРНТИ 14.01.85

**К ОЦЕНКЕ ЭФФЕКТИВНОСТИ САЙТА «МОНИТОРИНГ
ОБУЧАЮЩИХСЯ В ВУЗЕ»**

**САЙТТЫҢ ТИІМДІЛІГІН БАҒАЛАУ «УНИВЕРСИТЕТТЕГІ
СТУДЕНТТЕРДІҢ МОНИТОРИНГІ»**

**TO ASSESS THE EFFECTIVENESS OF THE SITE
«MONITORING STUDENTS IN THE UNIVERSITY»**

В.З. КРУЧЕНЕЦКИЙ, Р.Н. ЖАГПАРОВ
В.З. КРУЧЕНЕЦКИЙ, Р.Н. ЖАГПАРОВ
V.Z. KRUCHENECKI, R.N. ZHAGPAROV

(Алматинский технологический университет, Алматы, Казахстан)
(Алматы технологиялық университеті, Алматы, Қазақстан)
(Almaty Technological University, Almaty, Kazakhstan)
E-mail: ruslan-zhagparov@yandex.ru

Рассмотрены вопросы оптимизации показателей качества веб-сайта «Мониторинг обучающихся в ВУЗе». Исследована эффективность сайта, на основе оценки его времени загрузки. Выполнена проверка гипотезы о нормальном распределении эмпирически найденного закона. Определены значения времени загрузки – минимального, статистически значимых, диапазонов изменений, что является инновационным и практически значимым для повышения эффективности работы сайта.

Университеттегі студенттерді бақылауға арналған сайттың тиімділігі оның жүктеу уақытын бағалау негізінде зерттеледі. Эмпирикалық заңның қалыпты бөлу туралы гипотезасы сыналды. Жүктеу уақытының мәндері анықталатын - ең аз, статистикалық маңызды, өзгерістердің ауқымы, ол сайттың тиімділігін арттыру үшін инновациялық және іс жүзінде маңызды болып табылады.

The effectiveness of the site, designed for monitoring students in the university, is investigated based on an assessment of its download time. The hypothesis of the normal distribution of an empirically found law is tested. The values of the load time are determined - minimal, statistically significant, ranges of changes, which is innovative and practically significant for increasing the efficiency of the site.

Ключевые слова: сайт, мониторинг, математическое ожидание, дисперсия, стандартное отклонение, мода, медиана, квартиль, квантиль.

Негізгі сөздер: сайт, мониторинг, күту, ауытқу, стандартты ауытқу, режим, медиа, квартиль, квантиль.

Keywords: site, monitoring, expectation, variance, standard deviation, mode, median, quartile, quantile.

Введение

Известно, что мониторинг обучающихся в ВУЗе, включающий отслеживание таких важнейших показателей, как сведения по ним,

успеваемости, условия проживания, оплаты и др., является важной функцией управления процессом и качеством подготовки специалистов. значимой частью социально-экономи-

ческой эффективности. Поэтому создание сайта, отражающего мониторинг обучающихся, размещенного на соответствующем портале ВУЗа, является актуальной задачей.

Также известно, что эффективность сайта включает в себя такие основные части, как техническая, социально-экономическая, которые, естественно, взаимно связаны. В данной работе, в силу ограничений по объему статьи, основное внимание посвящено первой из них.

Эффективность технической составляющей, в первую очередь, зависит от аппаратно-программной реализации сайта, следствием которой является один из таких важнейших показателей, как время доступа к сайту. То есть, на качественные, экономические и другие показатели Web-сайтов в значительной мере влияет их время загрузки, которое при условии использования равноценных компьютерных средств, во многом зависит от конструкции Web-сайта, текущего трафика.

Исследования ряда разработанных Web-сайтов, и, в частности, сайта, предназначенного для мониторинга обучающихся в АТУ, его имитационной математической модели, показали, что среднее время загрузки одного из них составило 7с., его стандартное отклонение – 2с.; размах загрузок - 12с. Из общего числа загрузок, а их генеральные совокупности составляли после обоснования объема выборки подобно, описанным в [1, 4], не менее, чем по 60 ÷ 100 наблюдений, 95% колебалось в интервале от 2 до 11 с, а 66% - в диапазоне 5 – 9 с. Измерения времени загрузки проводились с помощью больших электронных часов с шагом 1 с.; погрешность измерений не превышала десятые доли с.

Чтобы построить кривую распределения времени загрузок, наблюдения сначала были статистически упорядочены, рандомизированы. Они откладывались по горизонтальной оси, а значения числа наблюдений – по вертикальной оси (рис. 1).

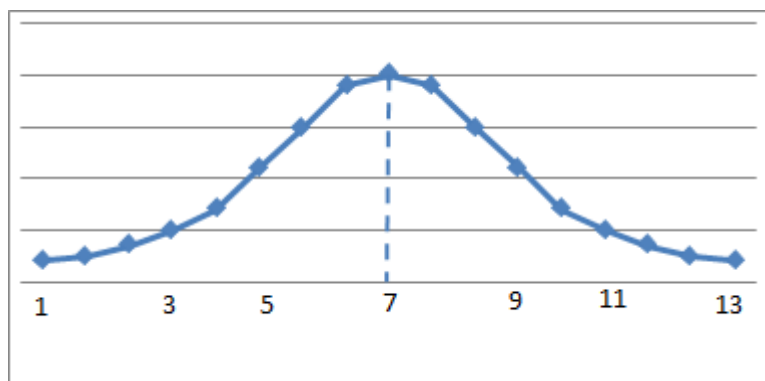


Рисунок 1 - Время загрузки Web – страницы сайта №1.

Исследуемая форма распределения времени загрузок оказалась близкой к кривой Гаусса, т.е. приняла нормальный закон распределения. Создание затем упорядоченного ряда данных позволило найти соответствующие статистические показатели, в том числе вычислить квантили стандартизованного нормального распределения. Далее была проведена всесторонняя проверка гипотезы о нормальном распределении времени загрузки Web-страницы.

Объекты и методы исследования

Объектом исследования является сайт, отражающий мониторинг обучающихся, в частности, его время загрузки. В числе основных методов исследования использованы

методы математической статистики, имитационные математические модели, аналитическая и графическая интерпретации результатов исследования.

Результаты, их обсуждение

Для оценки оптимальности конструкции Web-сайтов небезынтересно на основе полученной информации определить частоту времени загрузки Web-страницы, не превышающей определенное приемлемое время, и в каком интервале это время составит подавляющее большинство, например, - 99%, Обратимся к методам статистики. Время загрузки сайта является случайной величиной, причем численно – непрерывной, т.к. это время измеряется, а не подсчитывается. Математическим

выражением, описывающим распределение таких величин, является плотность непрерывного распределения их вероятности. В общем случае такое распределение может быть нормальным, равномерным, экспоненциальным или иным. Как указано выше, исследуемое нами распределение времени загрузки эмпирически оказалось нормальным, ибо оно подтверждает его основные свойства:

- имеет колоколообразную форму кривой;
- его математическое ожидание, мода и медиана близко совпадают друг с другом;

- основная масса нормально распределенных значений времени загрузки лежит в интервале, длина которого равна $\pm 1,33$ стандартного отклонения, т.е. межквартильный размах находится в интервале $\pm 0,66$ стандартного отклонения от среднего значения;

- значения нормально распределенной случайной величины лежат на всей числовой оси (теоретически, в пределах $-\infty \leq X \leq +\infty$).

При этом полагаем известным, что:

1. Математическое ожидание равно сумме всех значений генеральной совокупности измеренных величин загрузок, деленное на ее объем

$$\mu = \sum_{i=1}^N (x_i) / N, \quad (1)$$

где: μ - математическое ожидание, X_i - i -е наблюдение переменной X времени загрузки сайта, N - число наблюдений.

2. Стандартное отклонение δ , позволяющее оценить величину колебаний выборки вокруг среднего значения, равно корню квадратному из дисперсии

$$\delta = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}} \quad (2)$$

3. Медиана – число, разделяющее выборку пополам, т.е. 50% элементов меньше медианы, а 50%, - больше ее.

4. Квартиль разбивает упорядоченный набор данных на четыре части. Соответственно первый квартиль – это число, разделяющее его на две части: 25% меньше и 75% больше его; третий квартиль, наоборот - 75% больше, 25% - меньше его; межквартильный размах составляет 25% данных.

Плотность распределения вероятности (плотность нормального распределения) $f(x)$ не трудно вычислить по формуле

$$f(x) = \frac{1}{\sqrt{2\pi\delta}} * e^{-\frac{1}{2}(\frac{x-\mu}{\delta})^2}, \quad (3)$$

где: e – константа, равная 2,71828, π – константа, равная 3,14159.

Так как в формуле (3) e и π – математические константы, то плотность распределения зависит только от двух параметров μ и δ . Эти параметры часто называют моментами, соответственно первым и вторым. Плотности нормального распределения времени загрузки для Web-страниц одного и того же сайта могут иметь одинаковые первые моменты, но разные – вторые, или наоборот – разные первые и одинаковые – вторые, или и те, и другие разные. Поэтому в зависимости от комбинаций значений μ и δ . можно выделить характерные виды нормальных распределений, полученные для времени загрузки Web-страниц других исследованных сайтов (рис.2).

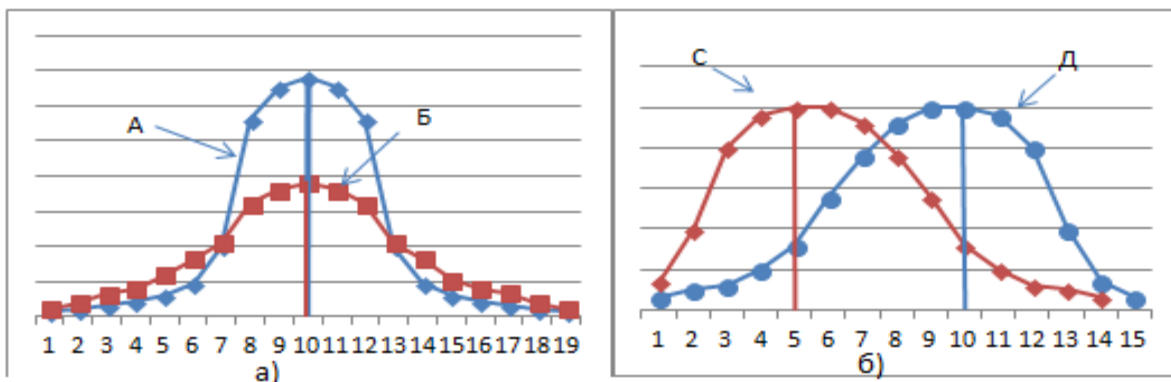


Рисунок 2 - Кривые нормального распределения. (а – симметричное, б – несимметричное с положительной и отрицательной симметрией).

На рис. 2 легко видеть, что распределения С и Д - несимметричные и характери-

зуются третьим моментом – коэффициентом асимметрии, кривые А и Б – симметричные,

но отличаются уплощением и характеризуются четвертым моментом – коэффициентом эксцесс. Для наиболее точной оценки симметричности кривых в статистике применяют пять базовых показателей: X_{\min} , X_{\max} , медиану, первый и третий квартили [2,3]. Для симметричного распределения: расстояния от X_{\min} или X_{\max} до медианы, от X_{\min} до первого квартиля и от X_{\max} до третьего квартиля, а также от этих квартилей до медианы – одинаковые.

Поскольку число комбинаций μ и δ велико, то вычислить значения плотностей вероятностей по формуле (3) достаточно сложно, поэтому разумно их преобразовать, нормируя любую нормально распределенную величину в виде Z , затем протабулировать. Для преобразования используем формулу

$$Z = \frac{X - \mu}{\delta} \quad (4)$$

Плотность стандартизованного нормального распределения находим, как

$$f(x) = \frac{1}{\sqrt{2\pi}} * e^{-\frac{1}{2}Z^2} \quad (5)$$

Любое множество нормально распределенных величин можно преобразовать в стандартизованную форму, а затем искомую вели-

чину вероятности определить по таблице кумулятивного стандартизованного нормального распределения, например, по таблице Д2, приведенной в [2] (стр. 1240).

Кривая Гаусса, построенная для генеральной совокупности по данным испытаний времени загрузки Web-сайта №1, как указано выше, имеет вид (рис 1); на ней $\mu = 7c.$, $\delta = 2c$. На шкале переменных этой зависимости X укажем найденные по формулам (1,2) соответствующие значения $\mu \pm \delta$, а также $\mu \pm 2\delta$, $\mu \pm 3\delta$. Им соответствуют значения μ , равные 1, 2, 3, 9, 11, 13. Далее по формуле (4) нанесем на шкале Z соответствующие значения; они равны: -3, -2, -1, 0, +1, +2, +3. (рис.3) Каждому значению переменной X соответствует нормированное значение Z , из чего следует, что время загрузки Web-страницы на одну единицу стандартного отклонения δ больше математического ожидания μ , Это очевидно, так как $Z = (9 - 7) / 2 = +1$. Соответственно, время загрузки, отличающееся на 3δ , равно -3 (т.к. $Z = (1 - 7) / 2 = -3$), и значит меньше μ . Таким образом, для времени загрузки Web-страницы стандартное отклонение можно считать его единицей измерения.

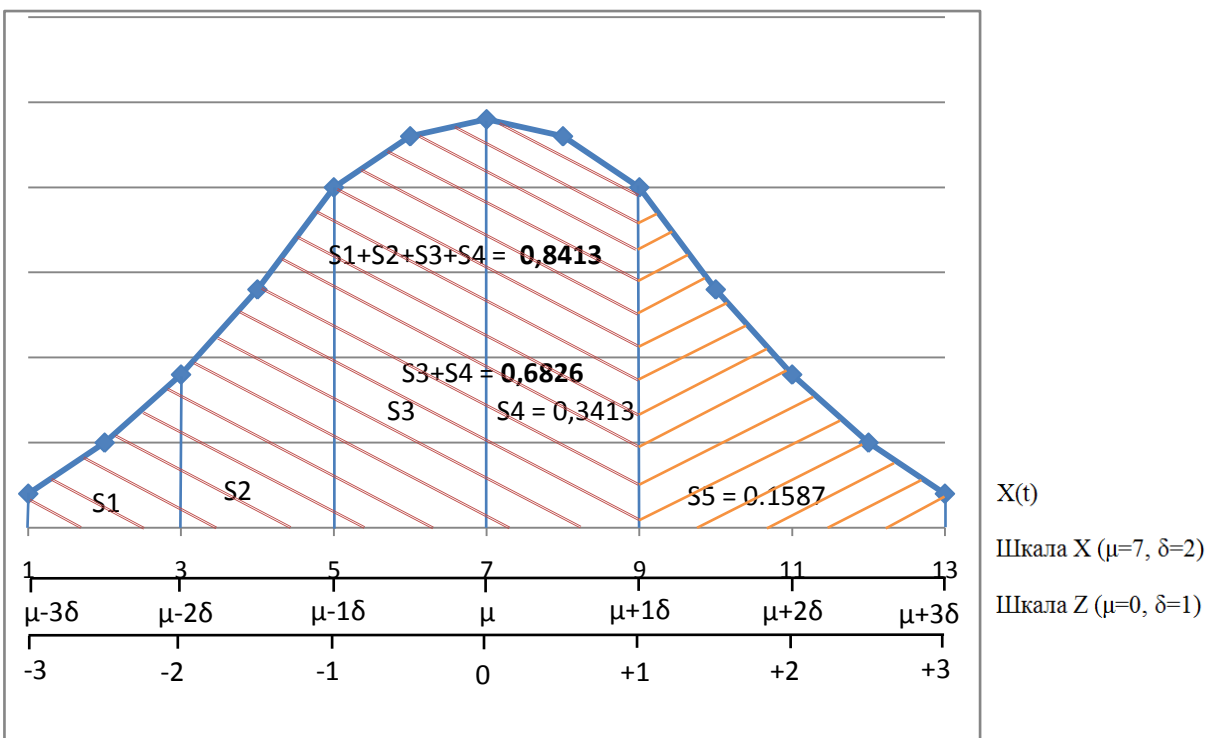


Рисунок 3 - Площадь фигуры, ограниченная интегральной кривой стандартизованного нормального распределения.

На рис.3 показан полигон относительных частот, соответствующих времени загрузки одного сайта. Поскольку результаты измерений образуют полную генеральную совокупность вероятностей, сумма вероятностей, т.е. площадь фигуры под кривой, должна быть равной единице.

Вернемся к нашей задаче определения вероятности того, что время загрузки Web-страницы не превышает определенное, например, 9с. Поскольку времени загрузки соответствует математическое ожидание, отличающееся на одно стандартное отклонение, то следует определить вероятность того, что

время загрузки не превышает величину $(\mu + \delta)$. Воспользуемся указанной выше таблицей Д2 [2] кумулятивной вероятности, т.е. площади фигур, ограниченных стандартизированной гауссовой кривой, и лежащих левее величины X , а по отношению к формуле (4) – левее Z . Для наглядности покажем фрагмент этой таблицы (табл. 1), ограничившись лишь некоторыми значениями Z из таблицы Д2, используемых для дальнейших вычислений в наших исследованиях. Заметим, что обычно Z записывается с двумя цифрами после запятой (в нашем примере Z записывается как +1,00).

Таблица 1 - Площади фигур, ограниченные кривой нормального распределения.

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5209	0.5239	0.5279	0.5319	0.4359
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.4	0.6554	0.6591	0.6628	0.5910	0.6664	0.6736	0.6772	0.6808	0.6884	0.6889
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7518	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8022	0.8054	0.8076	0.8106	0.5133
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8577	0.8599	0.8106	0.8621
.....										
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0259	0.0244	0.0239	0.0232
-1.7	0.0446	0.0436	0.0437	0.0418	0.0409	0.0401	0.0392	0.0384	0.0275	0.0267
-1.6	0.0549	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0175	0.0465	0.0455
.....										
-1.2	0.1151	0.1112	0.0526	0.1075	0.1056	0.0495	0.1038	0.2031	0.1003	0.0985

Чтобы найти вероятность, значения $Z = +1,00$, необходимо просмотреть столбец Z таблицы 1 и найти строку, соответствующую первым двум цифрам Z . Такой строкой является та, в которой стоит значение Z , равное 1,0. Далее переместимся по этой строке и найдем значение, находящееся на пересечении со столбцом, соответствующим сотым долям величины Z ; в нашем случае она равна 0, т.е. находится в первом столбце (0,00). Следовательно, искомая вероятность, находящаяся на пересечении соответствующих строки и столбца, равна 0,8413 и это означает, что Web-страница загружается меньше, чем за 9с и ее вероятность времени загрузки составляет 84,13%. Изложенное выше проиллюстрируем (рис.3), как время загрузки Web-страницы, определяемое площадью фигуры, ограниченной интегральной кривой стандартизованного распределения.

Нетрудно определить вероятность того, что время загрузки рассмотренной Web-страницы превысит 9с., как разность $1 - 0,8413 = 0.1587$. Также просто определить

вероятность того, что время загрузки Web-страницы лежит в определенном интервале, например, от 7 до 9с. Для этого необходимо определить, что загрузка продлится более 7с и вычесть из уже известной вероятности того, что она составляет не более 9с. Поскольку математическое ожидание и медиана нормального распределения совпадают, то вероятность превышения загрузки 7с. из (4) следует $Z = (7 - 7) / 2 = 0,00$. По таблице 1 определяем, что площадь фигуры, ограниченной нормальной кривой и значением $Z=0,00$, равна 0,05. Следовательно, площадь, лежащая под нормальной кривой между значением $Z = 0,00$ и $Z=-1,00$, равна $0,8413 - 0,5000 = 0,3413$.

Аналогично определяется из предыдущего случая вероятность того, что время загрузки Web-страницы лежит в интервале меньше 7с и больше 9с, поскольку является противоположным событием, т.е. $1 - 0,3413 = 0,6587$. Эта же оценка может быть сделана иначе, исходя из того, что вероятности событий $P(X<7)$ и $P(X>9)$ известны; сложим их и имеем $0,5000 + 0,1587 = 0,6587$.

Также просто определить и вероятность того, что время загрузки лежит в пределах от 5 до 9с., т.е. $P(5 < X < 9)$. Для этого, учитывая, что формула преобразования (4) позволяет вычислить лишь вероятность того, что случайная величина меньше определенного значения μ , а в данном случае интервал времени загрузки лежит по разные стороны математического ожидания, используем следующий алгоритм решения: вначале вычислим вероятности $P(X < 9)$ и $P(X < 5)$, затем вычтем результат один из другого. Первый результат нам уже известен, т.е. когда время загрузки не превышает 9с и равно 0,8413, а чтобы найти второй результат, вычислим по (4) значение $Z = (5 - 7) / 2 = -1,00$. Далее из таблицы 1 следует, что $Z = -1,00$ соответствует вероятность 0,1587. Тогда, окончательный результат равен $0,8413 - 0,1587 = 0,6826$.

Таблица 2 - Показатели погрешности результатов измерений времени загрузки

Интервал	Правило Бьенами-Чебышева	Эмпирическое правило
$(\mu - \delta; \mu + \delta)$	Минимум 0%	Приблизительно 68%
$(\mu - 2\delta; \mu + 2\delta)$	Минимум 75%	Приблизительно 95%
$(\mu - 3\delta; \mu + 3\delta)$	Минимум 88,89%	Приблизительно 99,7%

В нашем случае, при условии соблюдения правила 1δ , указанная вероятность составила 68,26,%; правила 2δ - несколько больше, чем 95% (95,44 %), а 3δ - 99.73%. Это значит, что для рассматриваемого сайта время загрузки Web-страницы, при условии соблюдения 1δ , лежит в пределах от 5 до 9 с.: 2δ - от 3 до 11 с, а при 3δ - в интервале от 1 до 13с.

Выше мы рассмотрели случаи определения вероятности времени загрузки Web-страницы в определенном интервале. Небезынтересно, пользуясь формулой (4) и таблицей 1, решить «обратную» задачу – нахождения значений X , соответствующих заданной интегральной вероятности. Воспользуемся следующим алгоритмом:

- построим гауссову кривую и отложим на ней математическое ожидание по шкале переменных X и Z . При этом используем выражение

$$X = \mu + Z\delta \quad (6)$$

Его легко найти из (5), т.к. $Z = (X - \mu) / \delta$, $Z * \delta = X - \mu$, $X = Z * \delta + \mu$. Затем:

- вычислим суммарную площадь фигуры, ограниченной нормальной кривой и лежащей слева от значения X ;

Таким образом, воспользовавшись формулой (4) и таблицей 1 (в общем случае $D2$ [2]), можно определить, что площадь S , расположенная под нормальной кривой и ограниченная значениями X и соответственно Z , равна:

- при $X=3$, $Z = -2,00$ $S = 0,0228$;
- при $X=1$, $Z = -3,00$ $S = 0,0135$;
- при $X=11$, $Z = +2,00$ $S = 0,9772$;
- при $X=13$, $Z = +3,00$ $S = 0,9986$.

Полученные результаты хорошо согласуются с правилом Бьенами-Чебышева и эмпирическими правилами, которые, как известно [2,3], определяют число данных, лежащих вокруг математического ожидания и говорят, что процент данных, находящихся в интервале, составляет соответствующий табл. 2.

- используя таблицу 1, определим Z , соответствующую найденной площади;

- с помощью формулы (6) вычислим значение X .

Например, чтобы найти, каково время загрузки Web-страницы при интегральной вероятности 0,1, значение переменной X для рассматриваемого сайта на рисунке 1 соответствует $X = 7 - 1,28 * 2 = 4,44$ (строке таблицы 1 соответствует значение 1,2, а столбцу - 0,08) То есть, в 10% случаев Web-страница будет загружаться быстрее, чем 4,44 с.

Аналогичные исследования были выполнены и для других вариантов сайта. Из которых следует, что для построенной кривой распределения одного из них оказалось математическое ожидание равно 4с, а стандартное отклонение - 1с. В этом случае для задачи определения вероятности времени загрузки Web-страницы, равной 5с, оказалось такое же значение, как и для предыдущего варианта сайта - 84,13%.

Еще одним подходом к проверке гипотезы о нормальном распределении, причем, весьма наглядным, является построение линии смещения с использованием квартилей, децилей или графика нормального распреде-

ления. Напомним, что децили разбивают изменение диапазона данных на десятые доли, а в процентах – на сотые доли. На степень близости к нормальному распределению будет указывать построенная прямая линия. Для

рассмотренного сайта такая линия показана на рис. 4. Что касается построения графика нормального распределения, то он строится аналогично, но по горизонтальной оси откладываются не X, а нормированные Z-значения.

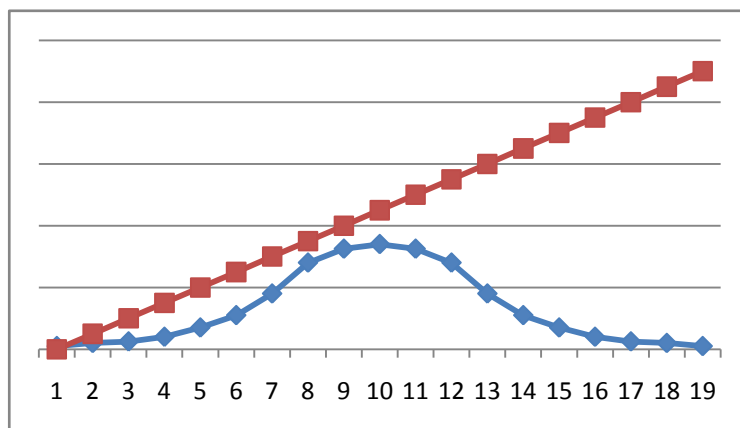


Рисунок 4 - Проверка гипотезы о нормальном распределении

Заключение

Выше отмечалось, что цель оценки времени загрузки Web–страницы, как одного из показателей его эффективности, в первую очередь связана с оптимизацией конструкции сайта. Последняя, напрямую зависит от требований, предъявляемых к функциональной и обеспечивающей части сайта. Требования к составу, содержанию сайта «Мониторинг обучающихся ВУЗа» вытекают из логической схемы, основу которой составляют входные и выходные данные сайта.

Каждая функционально законченная позиция в данной структуре реализуется на сайте с разной степенью детализации, раскрывая их содержание, предоставляя сведения об обучающихся, месте, условиях их проживания, оплаты и т.д.

Требования к обеспечивающей части сайта зависят от его аппаратной и программной реализации. В плане аппаратной части мы ориентировались на «обычные» общедоступные коммерческие компьютеры. Поэтому на время загрузки Web–страницы решающее влияние оказывает его программное обеспечение. В качестве последнего возможно использование различных программных средств. В данной работе при создании различных вариантов сайтов проанализированы особенности языков программирования: PHP, DHTML, SGML, XHTML, XML, Perl, Python. Из них наиболее простым является язык

HTML, его усовершенствованная версия XML. Однако, использование этих языков ориентировано на статические сайты и достаточно трудоемко. Использование PHP удовлетворяет всем требованиям динамических сайтов, но для расширения возможности по дизайну, защите требует дополнительного использования веб-сервера Apache, системы управления базами данных (СУБД) MySQL, системы управления содержимым сайта (CMS). В этой связи рассмотрены архитектура, основные функции, возможности, особенности использования систем управления созданием сайтов, их содержимым: Drupal, TYPO3, Mambo, WordPress, Joomla, проанализированы особенности современных обеспечивающих компонентов систем управления созданием сайтов: Денвер, WAMP, LAMP, серверов, СУБД: Apache, Apache HTTP, MySQL.

Выше, при оценке времени загрузки Web–страницы, рассматривалась проверка гипотезы о нормальном распределении на основании построенной экспериментальной кривой Гаусса и ее соответствие нормальному распределению на основе рассчитанных аналитически основных статистических характеристик. На практике характеристики набора данных могут несколько отличаться от теоретических. В таких случаях кривая может оказаться несимметричной; также может отличаться математическое ожидание и соответственно межквартильный размах не дол-

жен быть равным 1,33 стандартного отклонения. Поэтому для проверки гипотезы о нормальности можно применить другой подход, основанный на построении диаграмм и оценке их вида. В случае небольшого набора данных следует построить диаграмму «ствол и листья». Для больших наборов данных удобнее построить гистограмму или полигон. Затем, следует найти описательные статистики (моду, медиану, межквартильный размах и стандартное отклонение) и сравнить с теоретическими; после этого оценить распределение данных. Например, проверить лежат ли $2/3$ наблюдаемых данных в окрестности среднего значения на расстоянии не более одного стандартного отклонения или $4/5$ - на расстоянии не более $1,28$ его стандартного отклонения.

СПИСОК ЛИТЕРАТУРЫ

1. Крученецкий В.З, Акимкулова Д.С. О реализации технической компоненты электронного университета / Материалы Международной научно-практической конференции «Инновационное развитие пищевой, легкой промышленности, индустрии гостеприимства», посвященной 55-летию Алматинского технологического университета, 12-13 октября 2012 г. - С. 570-572.
2. Левин, Дэвид М., Стефан, Дэвид, Кребиль, Тимоти С., Беренсон, Марк Л. Статистика для менеджеров с использованием Microsoft Excel, 4-е изд. - М.: Издательский дом «Вильямс», 2014. - 1312 с.: ил.
3. Крученецкий В.З., Вязигин С.В., Отыншиев М.Б., Жилисбаева Р.О. К обоснованию объема выборки измерений тонины волокон шерсти с помощью электронного микроскопа // Известия ВУЗов. Сер. Технология текстильной промышленности. Ивановская государственная текстильная академия, № 5 (365). – 2016. - С. 68-72.