

ОЦЕНИВАНИЕ ПАРАМЕТРОВ КАЧЕСТВА ТЕСТОВОГО КОНТРОЛЯ В ОБРАЗОВАТЕЛЬНЫХ СИСТЕМАХ

Введение

Стабильность качества любого процесса в значительной мере зависит от выбранного вида и метода управления процессом и контроля его параметров.

Входной контроль знаний студентов проводится целью:

- определения целесообразности и степени корректировки подачи текущей дисциплины и необходимости возобновления знаний, умений и навыков по базовым дисциплинам,
- определения целесообразности корректировки учебных планов предыдущих дисциплин, в ходе которых получены проверяемые знания студентов.

На рис. 1 представлен фрагмент процессной модели образовательного цикла в вузе, детализирующий элемент входного контроля.

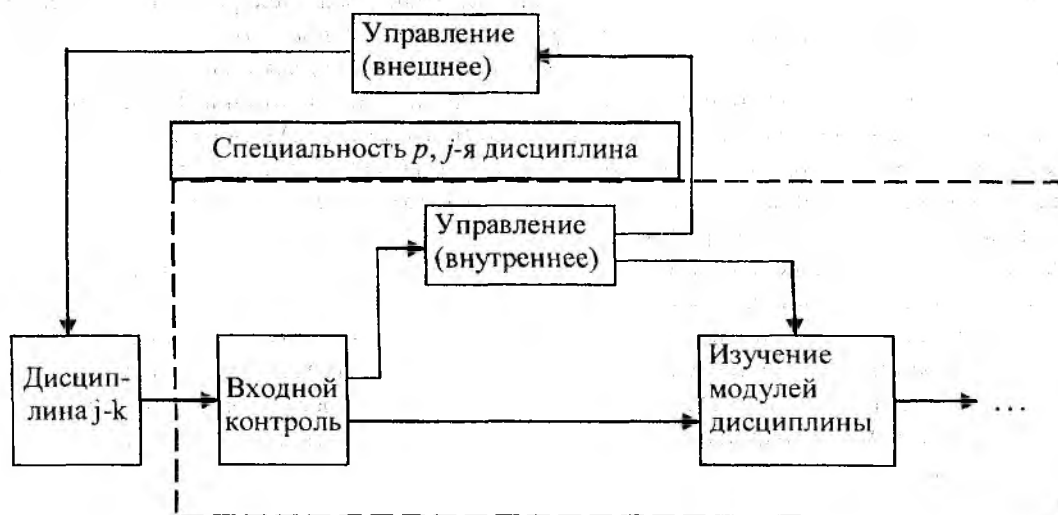


Рис. 1

При входном контроле знаний студентов группа должна в идеальном случае подвергаться сплошному контролю. На практике группа не всегда приходит в полном составе на входной контроль, следовательно, входной контроль получается, по сути, выборочным. Исходя из целей входного контроля знаний студентов, должен использоваться альтернативный выборочный контроль, когда все студенты в проверяемой группе разделяются на подгруппы – с достаточными и с недостаточными знаниями для успешного продолжения изучения курса без внесения корректив в рабочую программу, и последующее решение принимается в зависимости от соотношения количества студентов, оказавшихся в этих подгруппах.

При проведении входного контроля с целью принятия решений, относительно корректировки учебных планов, необходимо оценить степень достоверности этих решений. В этом смысле возникает необходимость ответить на следующие вопросы:

- каково минимальное число (из заданного объема группы) студентов должно пройти входное тестирование для обеспечения необходимого уровня достоверности;
- каково приемочное число не владеющих необходимым объемом базовых знаний студентов в выборке, позволяющее с достаточной долей уверенности принимать решение о необходимости изменения подачи материала в текущей дисциплине.

Основной показатель качества процесса входного контроля – достоверность – степень доверия к результатам контроля. В качестве показателя достоверности контроля используется вероятность принятия правильного решения по результатам контроля (соответствует основному принципу системы менеджмента качества – принятие решений на основе фактов).

Достоверность контроля является основной характеристикой системы контроля, оценивающей качество получаемой информации. Поэтому, при выборе количественной меры достоверности контроля, который может быть сделан различными способами, необходимо руководствоваться следующими соображениями. Достоверность контроля должна функционально зависеть от объема и точности измерения параметров, легко вычисляться и позволять достаточно просто решать обратную задачу – определять объем контроля и необходимую точность измерения параметров по заданной достоверности контроля.

Поскольку качество решений при контроле оценивается вероятностями ошибок, то достоверность контроля можно определить в следующем виде [1]:

$$D = 1 - p_{ош}, \quad (1)$$

где $p_{ош}$ – вероятность ошибочных решений при контроле, равная сумме рисков изготовителя и заказчика:

$$p_{ош} = \alpha + \beta, \quad (2)$$

следовательно

$$D = 1 - \alpha - \beta, \quad (3)$$

где α , β – ошибки 1, 2 рода.

Риск поставщика α : заведомо достаточные знания студентов ошибочно признаются за недостаточные. Учебные планы предыдущих курсов будут напрасно изменены, что повлечет финансовые и временные затраты.

Риск потребителя β : недостаточные знания студентов ошибочно принимаются за достаточные. Получится, что преподаватель не смог донести материал курса до студентов, и его могут признать некомпетентным.

Если отбросить личностные характеристики студентов и представить их как носителей знаний, умений и навыков, то при входном контроле знаний студентов можно применить методы статистического выборочного контроля по альтернативному признаку, аналогичные для технической продукции.

Обозначим следующие параметры:

- номинальное (списочное) количество студентов в группе – размер генеральной совокупности N ;
- количество студентов, пришедших на входной контроль – объем выборки n ;
- количество студентов в группе размером N , не владеющих необходимым объемом базовых знаний – количество брака в генеральной совокупности M ;
- доля студентов в группе размером N , не владеющих необходимым объемом базовых знаний – доля брака q в генеральной совокупности, $q = M/N$;
- максимальное число студентов в выборке, не владеющих необходимым объемом базовых знаний, являющееся критерием для принятия решения о том, что еще нет необходимости корректировать подачу материала – приемочное число A_c .

План контроля – совокупность требований и правил, которые следует соблюдать при решении о приемке партии продукции [2]. Вид плана для выборочного контроля по альтернативному признаку однозначно определяется как (N, n, A_c) [2, 3]. ДСТУ ISO 2859-1-2001 «Плани вибіркового контролю, визначені приймальним рівнем якості для послідовного контролю партій» в данном случае не позволяет выбрать план контроля. Существенными доводами этого являются следующие:

- маленький размер генеральной совокупности (около 30 человек);
- большой (относительно генеральной совокупности) объем выборки и случайность его размера;
- в отличие от входного контроля технической продукции, где риск поставщика α и риск потребителя β задаются непосредственно из экономических показателей и часто нормирова-

ны, при входном контроле знаний студентов значения α и β с затратами связаны косвенно, и, по сути, требуют определения.

Таким образом, целью исследования является определение плана входного контроля знаний студентов, при котором вероятность ошибки принятия неправильного решения ($\alpha + \beta$) будет маловероятной.

Определение плана входного контроля знаний студентов

Для любого плана входного контроля справедливы уравнения [2]:

$$\begin{aligned} P(q_0) &\geq 1 - \alpha, \\ P(q_m) &< \beta, \end{aligned} \quad (4)$$

где q_0, q_m – приемочный и браковочный уровни качества партии соответственно

Тогда, ошибка 1-го рода – это браковка партии с уровнем качества $q < q_0$ (хорошей). Ошибка 2-го рода – приемка партии с уровнем качества $q > q_m$ (не годной).

Вероятность того, что количество студентов m , не владеющих базовыми знаниями, в выборке не превышает приемочное число A_c , равна

$$P(q) = P_0 + P_1 + \dots + P_{A_c} = \sum_{m=0}^{A_c} P_m. \quad (5)$$

Вероятность того, что в выборке объемом n , взятой из генеральной совокупности объемом N , окажется ровно m студентов, не обладающих необходимым объемом знаний, находится из уравнения для гипергеометрического распределения (проверка гипотезы о виде закона распределения производилась помощью критерия согласия χ^2 [4]):

$$P_m = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}, \quad (6)$$

где $C_M^m = \frac{M!}{m!(M-m)!}$ – число сочетаний из M по m (остальные сочетания вычисляются аналогично).

Следовательно, уравнения плана контроля имеют вид:

$$1 - \alpha = \sum_{m=0}^{A_c} \frac{C_{M_0}^m C_{N-M_0}^{n-m}}{C_N^n} \quad (7)$$

$$\beta = \sum_{m=0}^{A_c} \frac{C_{M_m}^m C_{N-M_m}^{n-m}}{C_N^n} \quad (8)$$

где $M_0 = N \cdot q_0$, $M_m = N \cdot q_m$.

Для определения достоверности входного контроля (вероятности принятия правильного решения) необходимо решить следующие задачи:

1. Определить минимальное число студентов n , которые должны прийти на входной контроль.
2. Определить приемлемый уровень качества q_m .
3. Определить приемочное число студентов A_c .
4. Определить зависимость достоверности входного контроля от значения приемочного числа A_c для различного числа студентов n , пришедших на входной контроль.

Для того чтобы определить достоверность входного контроля (рис. 2) для каждой тестируемой группы преподаватель должен задать процент студентов в группе, владеющих достаточным уровнем знаний по базовым дисциплинам, который будет критерием принятия решения о корректировке рабочей программы. В проектируемой системе необходимо автоматически рассчитывать значение приемочного уровня качества q_0 и рекомендуемый объем выборки n' .

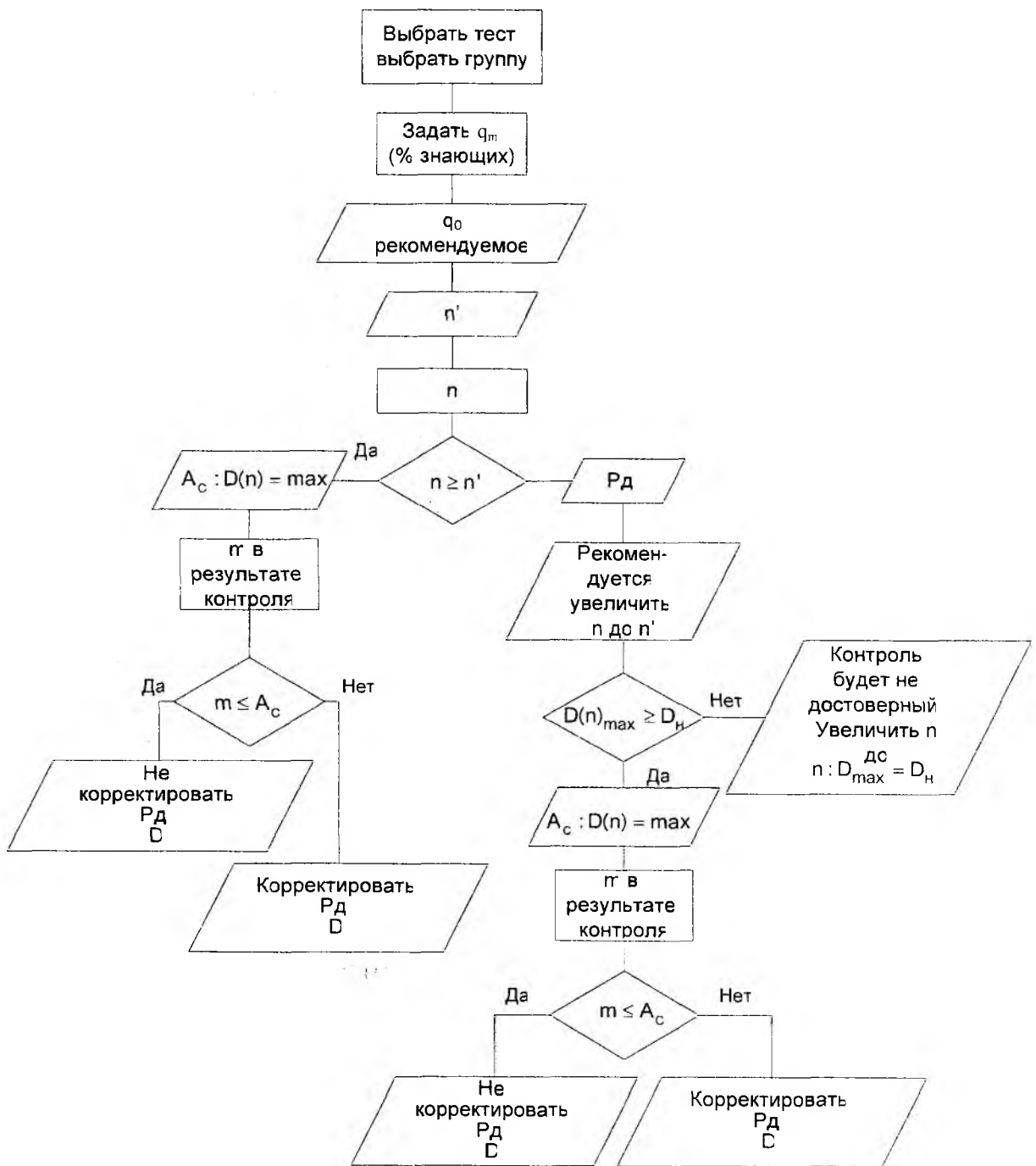


Рис. 2

Математическим ожидание доли не справившихся с тестами студентов в выборке: $M\left(\frac{m}{n}\right)$. Поэтому правило принятия решения при выборочном контроле должно было бы формулироваться следующим образом: при $M\left(\frac{m}{n}\right) \geq q_m$ необходима коррекция рабочей программы; при $M\left(\frac{m}{n}\right) < q_m$ коррекция не нужна. Однако определить $M\left(\frac{m}{n}\right)$ невозможно. Можно установить лишь долю бракованных изделий $\frac{m}{n}$ в конкретной выборке объемом n . Она отличается от $M\left(\frac{m}{n}\right)$ на $\varepsilon \leq t\sigma_{m/n}$ с некоторой доверительной вероятностью [2]. Реко-

мендуется задавать значение $P_{\alpha} = 0,9$. Т.о, рекомендуемый объем выборки n' обеспечит выборочную ошибку ε меньше либо равной заданной с доверительной вероятностью 0,9.

Реальный объем выборки n равен количеству зарегистрировавшихся в этом тесте студентов.

Если $n \geq n'$, то задается такое значение приемочного числа A_c , при котором значение достоверности D для данного объема выборки n принимает максимальное значение.

В результате тестирования определяется количество брака в группе m . Если $m \leq A_c$, то принимается решение о том, что нет необходимости корректировать подачу материала с доверительной вероятностью $P_{\alpha} = 0,9$ и достоверностью $D = D(n)_{max}$. В противном случае, принимается решение о необходимости корректировки подачу материала с доверительной вероятностью $P_{\alpha} = 0,9$ и достоверностью $D = D(n)_{min}$. При этом преподаватель может посмотреть, какие вопросы теста вызвали у тестируемых затруднения, и определить, какие именно модули (вопросы) надо повторять.

Если $n \leq n'$, то рассчитывается доверительная вероятность приемки P_{β} или забраковки группы для такого n и при этом выводится сообщение о том, что рекомендуется увеличить n до n' . Если $D(n)_{max} \leq D_n$, D_n – нормативное значение достоверности (рекомендуется задавать равным 0,8) то выводится сообщение «контроль будет не достоверный, увеличьте объем выборки». Если $D(n)_{max} \leq D_i$, то значение приемочного числа A_c задается аналогично, как и для случая с $n \geq n'$, только приниматься или браковаться группа будет с доверительной вероятностью P_{α} , рассчитанной для данного $n \geq n'$.

Приемочный уровень качества q_0 можно задать исходя из среднего входного уровня дефектности q_{cp} . Средний входной уровень дефектности – математическое ожидание значения уровня дефектности в нескольких партиях, поступающих на контроль, или потоке продукции за определенный интервал времени [2]. С этой целью проводят сплошной контроль определенного числа групп.

Оценку среднего входного уровня дефектности можно получить как отношение общего числа не справившихся с тестами студентов во всех проконтролированных группах к общему числу студентов, т.е.

$$q_{cp} = \frac{\sum_{M=0}^N M \cdot k_M}{\sum_{i=1}^k N_i} \quad (9)$$

где k_M – число групп с M не справившимися с тестами студентами.

Расчет q_{cp} должен производиться для однородных объектов: для различных специальностей, которые в свою очередь делятся на специальности с низким и высоким проходным баллом, для различных дисциплин – технических или основных, с высокой трудностью и гуманитарных или второстепенных для данной специальности, с низкой трудностью, для различных курсов – младших и старших.

Для получения среднего входного уровня дефектности использовались данные о сплошном входном контроле 200 групп 1–5 курсов 47 дисциплин Харьковского национального университета радиотехники за 2008 год. В табл. 1 представлен фрагмент результатов сплошного входного контроля, рассчитанные средние входные уровни дефектности и средние приемочные числа для указанных дисциплин.

Из таблицы видно, что даже на одном курсе, в одних и тех же группах получается большой разброс q_{cp} по разным дисциплинам. Потому, q_{cp} необходимо рассчитывать для каждой дисциплины отдельно.

		HTML	Почта	Интернет	Браузеры	Поиск	FTP	Σ
	<i>n</i>	22	23	20	24	24	23	136
КИ-08-1	<i>m</i>	2	1	1	1	3	0	8
	<i>n</i>	21	21	21	21	22	21	127
КИ-08-2	<i>m</i>	3	2	0	2	1	0	8
	<i>n</i>	24	22	23	26	23	23	141
КИ-08-3	<i>m</i>	9	6	6	4	7	5	37
	<i>n</i>	25	28	25	29	24	23	154
КИ-08-4	<i>m</i>	10	5	7	8	3	5	38
	<i>n</i>	20	20	20	21	20	20	121
КИ-08-5	<i>m</i>	3	1	0	1	6	0	11
	<i>n</i>	25	27	23	27	27	26	155
КИ-08-6	<i>m</i>	6	4	1	3	7	1	22
	<i>n</i>	137	141	132	148	140	136	834
Σ	<i>m</i>	33	19	15	19	27	11	124
$q_{\text{ср}}$		0,241	0,135	0,114	0,13	0,192	0,08	

Если правило принятия решения заменить на практически реализуемое – при $\frac{m}{n} \geq q_i$ требуется коррекция программы, при $\frac{m}{n} < q_i$ коррекция не нужна, то оно будет справедливо с вероятностью, соответствующей

$$t = \frac{\varepsilon}{\sigma_{\frac{m}{n}}} \quad (10)$$

Если эта вероятность задана ($P_0 = 0,9$) и установлена точность, с которой допускается замена $M\left(\frac{m}{n}\right)$ на $\frac{m}{n}$ (по умолчанию $\varepsilon = 0,1$), то из последнего выражения можно найти необходимый объем выборки n .

При входном контроле знаний студентов дисперсия доли студентов в выборке не справившихся с тестами соответствует:

$$D\left(\frac{m}{n}\right) = \frac{1}{n^2} D(m) = \frac{q(1-q)}{n} \cdot \frac{N-n}{N-1}. \quad (11)$$

Отсюда необходимый объем выборки находится из условия [2, 5]:

$$\varepsilon = t \sqrt{\frac{q(1-q) \cdot (N-n)}{n(N-1)}} \quad (12)$$

Объем выборки для конечной генеральной совокупности:

$$n = \frac{n_0 N}{n_0 + (N-1)} \quad (13)$$

где $n_0 = \frac{t^2 q(1-q)}{\varepsilon^2}$.

Входящая в формулу доля не справившихся с тестами студентов q в группе не может быть известна. Это предполагает необходимость использования некоторой априорной информации. Обычно исходят из того, какой эта доля была примерно в тех же условиях. Если доля признака в генеральной совокупности q заранее неизвестна, для определения объема выборки следует задать $q = 0,5$, чтобы исключить недооценку объема выборки, т.к. тогда достигается максимальное значение числителя.

Например, для того чтобы можно было охарактеризовать группу из 27 человек достаточно, чтоб на входной контроль явилось 20 человек. Предельная ошибка выборки с вероятностью 0,9 не будет превышать 0,1.

Зависимость достоверности входного контроля знаний студентов от значения приемочного числа A_c для различного числа студентов n , пришедших на входной контроль, представлена на рис. 3.

Например, достоверность (правильность) выбора $A_c = 6$ для объема выборки $n = 20$, который обеспечит $\epsilon \leq 0,1$ с $P_0 = 0,9$ равна 0,92.

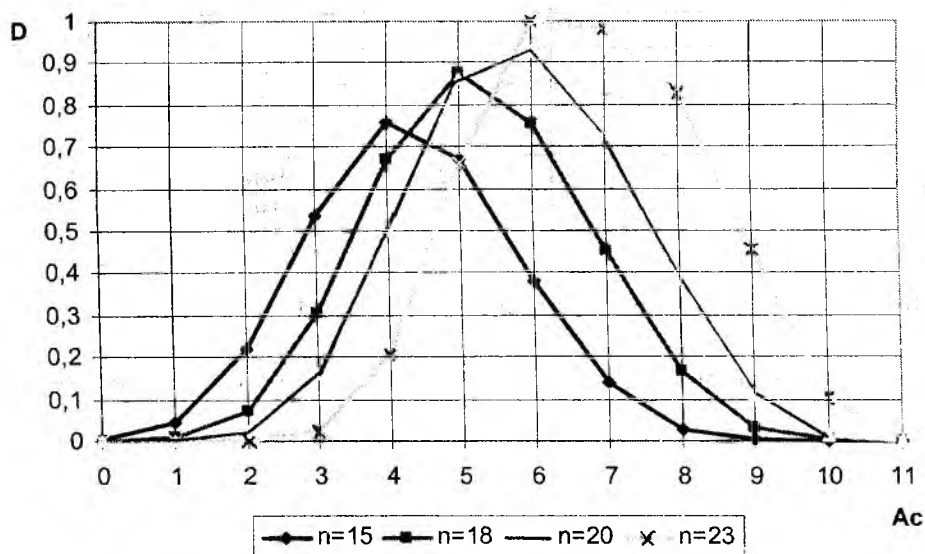


Рис. 3

Выводы

1. Разработанный метод позволяет правильно (с заданной вероятностью) реализовать критерий принятия решения при внутреннем управлении процессом образования с ограниченной выборкой партии тестируемых.

2. Данный метод оценки достоверности входного контроля знаний студентов является достаточно гибким и открытым к адаптации под новые условия.

3. Предложенный метод оценки достоверности входного контроля предполагает использование «идеальных тестов», как инструмента оценки знаний.

4. В дальнейшем предполагается исследование свойств и параметров тестов (трудность заданий, границы абсолютной успеваемости, время получения контролируемых (остаточных) знаний) для определения их влияния на достоверность выполнения предлагаемых критериев принятия решения, как по внутреннему, так и по внешнему управлению учебного процесса.

Список литературы: 1. Иванов Ю. П. Контроль и диагностика измерительно-вычислительных комплексов. СПб. СПбГУАП, 2004. 98 с. 2. Шишкин И. Ф. Прикладная метрология. М.: РИЦ «Татьянин день», 1993. 150 с. 3. ДСТУ ISO 2859-1-2001 Статистичний контроль. Вибірковий контроль за альтернативною ознакою. Частина 1. Плани вибіркового контролю, визначені приймальним рівнем якості для послідовного контролю партій. 4. Вадзинский Р. Статистические вычисления в среде Excel. Библиотека пользователя. СПб. Питер, 2008. 608 с. 5. Левин Д. М. Статистика для менеджеров с использованием Microsoft Excel. М. Вильямс, 2004. 1312 с.

Харьковский национальный
университет радиоэлектроники

Поступила в редколлегию 05.06.2009