

Лекция 2. Методы оценки качества программного обеспечения

2.1. Оценка качества программного обеспечения

На схеме (рис.2.1) представлена модель процесса оценивания, который применим для каждого компонента и системы в целом на любом этапе жизненного цикла ПП. Процесс предполагает последовательное выполнение **трех стадий**: 1) определение требований к качеству ПП, 2) подготовка к оцениванию и 3) оценивание.

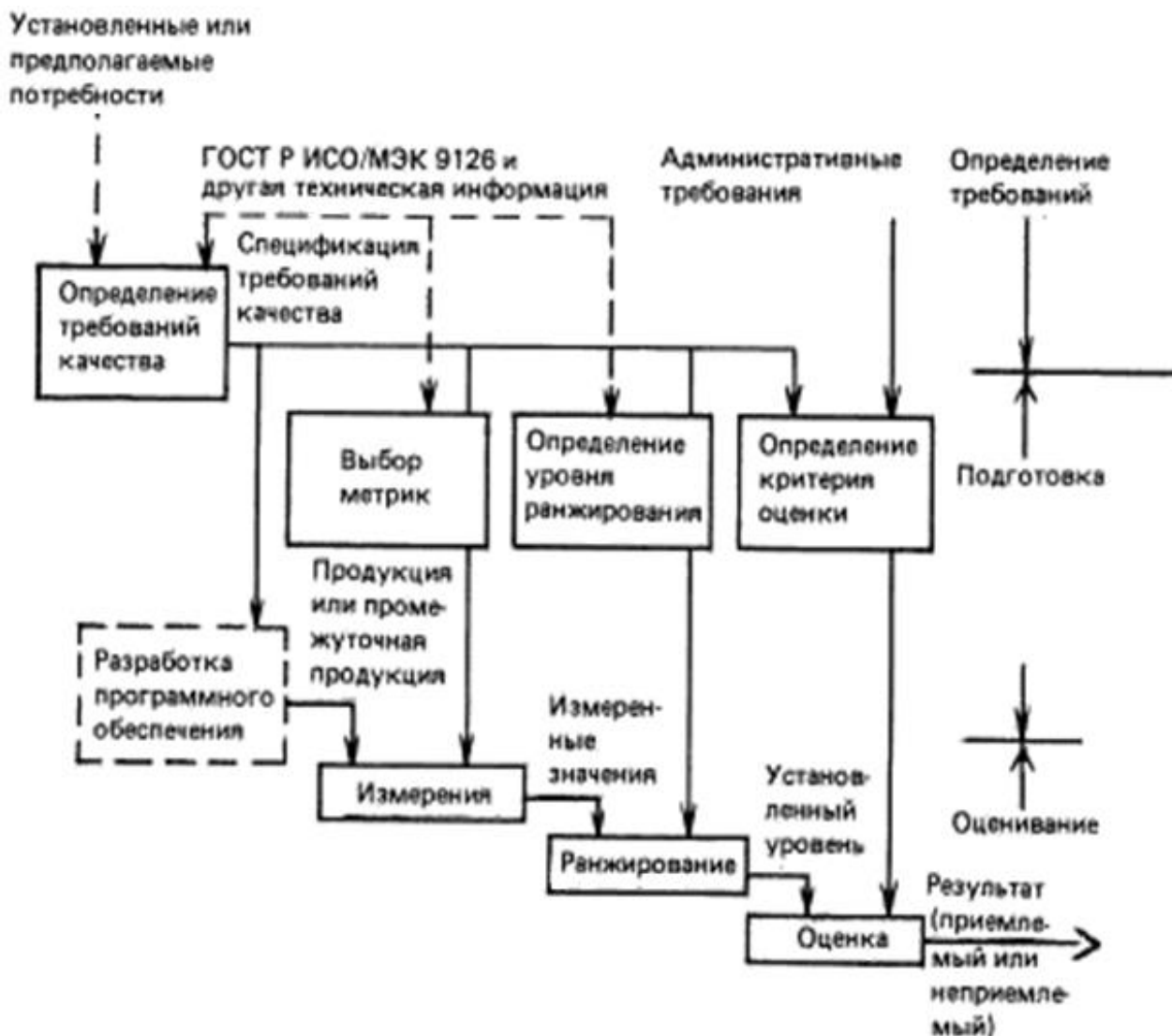


Рисунок 2.1 – Модель процесса оценивания ПП по ГОСТ 9126-93

1) **Установление требований к качеству ПП.** Требования к качеству ПП в целом и ее компонентов должны быть определены до начала разработки. Качественные характеристики ПП выражают посредством комплексных показателей и их подхарактеристик.

В различных нормативных документах номенклатура показателей качества ПП несколько отличается друг от друга. В **ГОСТ 28195-89** на верхнем уровне иерархической структуры расположено **шесть** характеристик: 1) надежность, 2) корректность, 3) удобство применения, 4) эффективность, 5) универсальность и 6) сопровождаемость. На втором уровне эти характеристики конкретизируются посредством **19-ти** комплексных показателей, а на третьем уровне оценочных

элементов уже более **двухсот**. Конкретные комплексные показатели и оценочные элементы для данного проекта выбирают в зависимости от назначения, функций и этапов жизненного цикла программного продукта. В документе ГОСТ Р ИСО/МЭК 9126) основных характеристик также **шесть**, но у половины из них другое название: 1) функциональная пригодность, 2) надежность, 3) применимость, 4) эффективность, 5) сопровождаемость и 6) переносимость. Для их детализации используется **21** показатель.

В таблице 2.1 представлены основные критерии, которые с минимальным дублированием описывают качество ПП, и примитивы, их детализирующие.

Таблица 2.1 – Критерии качества

Критерий	Примитивы качества
Функциональная пригодность	Пригодность для применения, уровень завершенности, точность, защищенность, способность к взаимодействию, согласованность со стандартами и правилами проектирования
Надежность	Уровень завершенности, работоспособность, стабильность, точность, автономность, устойчивость, восстанавливаемость
Удобство применения	П-документированность, информативность (документации по применению), коммуникабельность, устойчивость, защищенность
Эффективность	Временная эффективность, ресурсоемкость, характер изменения ресурсов, уровень автоматизации, эффективность по ресурсам и по устройствам, характер изменения во времени
Сопровождаемость	Анализируемость, изменяемость, стабильность, тестируемость
Мобильность	Адаптируемость, структурированность, замещаемость и внедряемость

Ниже для примера даны определения некоторых примитивов качества:

- Завершенность – степень обладания ПП всеми необходимыми частями и чертами, требующимися для выполнения его явных и неявных функций.
- Пригодность – наличие и соответствие набора функций конкретным задачам.
- Точность – приемлемость величины погрешности в выдаваемых программами ПП результатах с точки зрения предполагаемого их использования.
- Автономность – способность ПП выполнять предписанные функции без помощи или поддержки других компонент программного обеспечения.
- Устойчивость – способность ПП продолжать корректное функционирование, несмотря на задание неправильных (ошибочных) входных данных.
- Защищенность – способность ПП противостоять преднамеренным или нечаянным деструктивным (разрушающим) действиям пользователя.
- П-документированность – наличие, полнота, понятность, доступность и наглядность учебной, инструктивной и справочной документации, необходимой для применения ПП.

Стандарты не устанавливают содержания комплексных показателей качества (но приводят примеры их описания). Также стандарт не определяет методы измерения, ранжирования и оценки показателей качества. **Требования качества** для ПП в целом определяются совместно заказчиками и разработчиками при составлении технического задания и **фиксируются в спецификации качества** ПП.

Названия и содержание показателей качества для конкретного проекта может отличаться от вышеперечисленных. Например, в требованиях могут быть указаны такие характеристики, как расширяемость, многократность, совместимость, производительность.

Расширяемость – означает возможность развития системы, добавления отдельных элементов, внесения изменений. Расширяемость можно рассматривать как меру количества усилий, необходимых для реализации расширения.

Многократность – повторное использование кода, применяется для сокращения трудозатрат при разработке сложных систем. В новый проект включаем правильные, проверенные, опробованные в предыдущих ПП компоненты, качество которых уже является высоким. Сэкономленное время и энергия могут быть направлены на улучшение других характеристик качества ПП.

Совместимость – способность различных объектов (аппаратных или программных компонентов) взаимодействовать друг с другом, возможность использования ПП с другим программным обеспечением

Производительность характеризуется временем выполнения заданных транзакций или длительных операций.

2) Подготовка к оцениванию. Включает *выбор метрик* (показателей) качества, определение уровней ранжирования и критериев оценки этих метрик.

Для каждой указанной в задании пользовательской характеристики качества программной продукции подбирают метрики, используемые в процессе разработки. В качестве метрики могут быть приняты любой количественный признак и каждое количественно оцениваемое взаимодействие программного обеспечения с его окружением.

Можно выделить три **вида метрик** качества ПО:

А) *Определяющие наличие или отсутствие признака*: наличие гибкости, простоты освоения, структурированности.

Б) *Ранжируемые* метрики. Характеристики, которые можно ранжировать (расположить в порядке возрастания или убывания), сравнивая их с опорными. При этом указывается приоритетность признаков. Например, уровень завершенности может быть высоким, средним, низким.

В). *Числовые* параметры: время наработки на отказ, вероятность ошибки, число маршрутов в программе, т.д.

Для определения уровня ранжирования необходимо выбрать или построить измерительную шкалу. При оценке качества используются все **виды измерительных шкал**:

- номинальная – набор категорий. Можно выявить принадлежность характеристики какому-либо классу, применяется для метрик первой группы;
- порядка – упорядоченный по возрастанию или убыванию набор делений шкалы (опорных точек). Применяется для метрик второй группы;
- интервалов – упорядоченная шкала с равноудаленными делениями и произвольно выбранным нулём;
- отношений – упорядоченная шкала с равноудаленными делениями и естественным нулём. Для измерения числовых параметров применяются шкалы интервалов и отношений.

Полученное в ходе измерения значение метрики качества ничего не говорит о достаточности (или недостаточности) этого значения. Необходимо сравнение с нормой. Как минимум, нужно выделить значение на шкале, выше которого уровень качества считается удовлетворительным. А можно разделить шкалу на диапазоны,

соответствующие различному уровню удовлетворения требований (рис. 2.2). *Уровни ранжирования* следует определять для каждого конкретного показателя качества ПП. В числе *критериев оценки* время, стоимость, средние взвешенные, таблицы решений и другие.

3) **Процедура оценивания** включает измерение, ранжирование и оценку показателя качества. В ходе инструментальных или безинструментальных *измерений* определяют значения выбранных метрик. В ряде случаев сбор измерительной информации осуществляется путем регистрации событий. В случае косвенных измерений, обработки статистической информации и т.п. для получения значения метрики могут потребоваться расчеты. Далее определяют *ранг* для измеренного значения и на его основе делают *выводы* о достигнутом качестве программного продукта, принимают решение о дальнейшей судьбе ПП (приемка и выпуск, доработка, отбраковка).

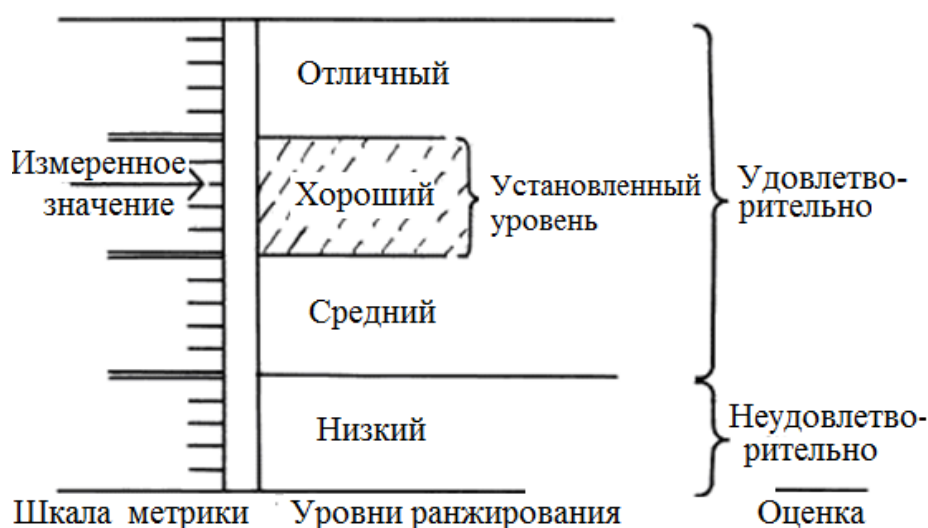


Рисунок 2.2 – Уровни ранжирования

2.2. Методы оценки качества программных средств

Измерения показателей качества могут быть 1) инструментальными и 2) безинструментальными. Некоторые метрики получают так называемым 3) регистрационным методом.

1) В большинстве случаев предпочтение следует отдавать **инструментальным** измерениям. Однако, список показателей качества, для которых существуют средства измерения, весьма ограничен. Например, в их числе время наработки на отказ, показатели производительности (потребление ресурсов центрального процессора, потребление оперативной памяти, время выполнения запроса), потребление сетевых ресурсов, время отклика.

2) При **безинструментальных** измерениях роль «нетехнического» средства измерений играет человек (эксперт). **Эксперт** – грамотный, высококвалифицированный, компетентный в рассматриваемых вопросах и достаточно опытный специалист, приглашаемый для выдачи квалифицированного заключения или суждения. Экспертная оценка выставляется по итогам изучения программ и документации на ПП, в ходе которого эксперт составляет свое мнение об обладании программным продуктом требуемым примитивом качества. **Экспертные методы** оценивания показателей качества применяют тогда, когда

использование технических средств измерения, невозможно, сложно или экономически не оправдано.

Повысить достоверность, точность, надежность и воспроизводимость экспертных оценок позволяет привлечение группы экспертов (**экспертной комиссии**). Точность средних оценок зависит от числа и компетентности экспертов, входящих в группу, согласованности их мнений. Обычно число экспертов в группе от семи до двадцати человек (и более при **социологических исследованиях**, например, при опросе потребителей продукции).

Эксперты проходят специальное предварительное обучение и инструктаж. Целесообразно протестировать экспертов, выявить их самооценку и взаимооценку (при наличии опыта совместной работы). Самооценку определяют путем анкетирования. При взаимооценке эксперты оценивают друг друга (в баллах). Компетентность и профпригодность экспертов устанавливают тестированием (им предлагается решить одну или несколько задач, подобных реальным, с известными ответами).

Важно также оценить надежность каждого эксперта и согласованность мнений членов группы. Степень надежности эксперта $H_{\text{э}} = \text{Эс} / \text{Эобщ}$, где Эс – число случаев, когда мнение эксперта совпало с результатами экспертизы, Эобщ – общее число экспертиз, в которых он участвовал. Для эксперта, работающего в группе, также вводят понятие об его относительной надежности $H_{\text{о}} = H_{\text{э}} / H_{\text{с}}$. Здесь $H_{\text{с}}$ – средняя степень надежности для данной группы экспертов.

Мерой согласованности мнений экспертов служит **коэффициент конкордации** Кендалла W , который определяют по формуле:

$$W = \frac{12S}{n^2(m^3 - m)};$$

где S — сумма квадратов отклонений суммы рангов каждого объекта экспертизы от среднего арифметического рангов; n — число экспертов; m — число объектов экспертизы. При отсутствии согласованности мнений экспертов коэффициент конкордации принимает значение нуль. При полном единодушии он равен единице. Если степень согласованности мнений не достаточна, принимают меры для ее повышения (проведение тренировок, обсуждение результатов и ошибок).

Экспертные методы основаны на использовании обобщенного опыта и интуиции специалистов. *Опрос эксперта* может проводиться путем анкетирования, интервьюирования, открытого или тайного голосования и т.п.

По способу проведения экспертизы, различают методы:

- непосредственной оценки;
- ранжирования;
- попарного сопоставления;
- последовательного сопоставления и некоторые другие.

а) Экспертный метод **«непосредственной оценки»** требует наличия измерительной шкалы, по которой значения показателей качества определяются сразу в установленных единицах (в баллах, нормо-часах, рублях, единицах условного топлива и т.д.). Балльные оценки даются экспертами непосредственно или получаются в результате формализации процесса оценки. Обычно используют пяти-, семи- или десятибалльную шкалы. Однако эта оценка может производиться и по двухбалльной системе: "обладает" или "не обладает". Обобщенный показатель качества находят как среднее арифметическое значение оценок, поставленных всеми экспертами.

б) Метод **ранжирования** предполагает расстановку объектов экспертизы в порядке их предпочтения, по важности или весомости. В случае работы экспертной группы на основе рядов отдельных экспертов строится обобщенный ранжированный ряд. Чем выше ранг (место, занятое при такой расстановке), тем предпочтительней объект. Порядок построения ранжированного ряда, следующий:

1. Все объекты оценки (изделия, свойства) нумеруются произвольно;
2. Эксперты ранжируют объекты по шкале порядка: численное значение ранга в ряду возрастающей шкалы порядка увеличивается от 1 до n (n — количество оцениваемых объектов);
3. Ранжированные ряды объектов, составленные экспертами, сопоставляются, определяются суммы рангов каждого из объектов экспертной оценки;
4. На основании полученных сумм рангов строят обобщенный ранжированный ряд.
5. Обобщенные экспертные оценки качества рассматриваемых объектов экспертизы, т.е. коэффициенты их весомости, рассчитываются по формуле:

$$g_i = \frac{\sum_{i=1}^n G_{i,j}}{\sum_{i=1, j=1}^{n,m} G_{i,j}},$$

где n — количество экспертов, m — число «взвешиваемых» показателей, $G_{i,j}$ — коэффициент весомости j-го показателя (ранг) в баллах, данный i-м экспертом. Коэффициенты весомости позволяют не только указать, какой объект лучше или хуже других, но и на сколько.

Пример Результаты ранжирования пятью экспертами семи объектов экспертизы следующие:

Первый эксперт: Q5<Q3<Q2<Q1<Q6<Q4<Q7;
 Второй эксперт: Q5<Q3<Q2<Q6<Q4<Q1<Q7;
 Третий эксперт: Q3<Q2<Q5<Q1<Q6<Q4<Q7;
 Четвертый эксперт: Q5<Q3<Q2<Q1<Q4<Q6<Q7;
 Пятый эксперт: Q5<Q3<Q1<Q2<Q6<Q4<Q7.

Построить обобщенный ранжированный ряд и определить весомость членов ряда.

Решение:

- 1) Сумма рангов: Q1 равна 4+6+4+4+3=21
 Q2 равна 3+3+2+3+4=15
 Q3 равна 2+2+1+2+2=9
 Q4 равна 6+5+6+5+6=28
 Q5 равна 1+1+3+1+1=7
 Q6 равна 5+4+5+6+5=25
 Q7 равна 7+7+7+7+7=35

Обобщенный ранжированный ряд Q5<Q3<Q2<Q1<Q6<Q4<Q7.

- 2) $\sum_{i=1, j=1}^{n,m} G_{i,j} = 21+15+9+21+7+25+35=140$

Весовые коэффициенты:

$g_1=21/140=0,15$; $g_2=15/140=0,11$; $g_3=9/140=0,06$; $g_4=28/140=0,2$;
 $g_5=7/140=0,05$; $g_6=25/140=0,18$; $g_7=35/140=0,25$;

Вывод: Лучшим по качеству признан продукт №5, на 0,01 отстает от него продукт №3.

в) Метод **парного сопоставления**. Парное сопоставление лежит в основе любого выбора. Сравнить две величины между собой легче, чем сразу определить их место на шкале порядка. Особенно, если объектов экспертизы сравнительно много. Измеряемые размеры сравнивают друг с другом, а результаты сравнения заносятся в таблицу, где для каждой пары указывают номер объекта, которому следует отдать предпочтение. Лучшим считается объект, чей номер повторился наибольшее количество раз. Иногда учитывается равноценность. Ранжирование производится на основе результатов парного сопоставления.

Порядок парного сопоставления следующий:

1. Составляется таблица, в которую эксперт заносит результаты сопоставления рассматриваемых объектов. При этом каждый i -й объект сопоставляется с другими j -ми объектами сравнения. Если i -й объект признается качественнее j -го, то это обозначается цифрой 1, противоположная оценка обозначается -1 , а равнокачественные объекты отмечаются в таблице цифрой 0 (ноль);

2. Данные о предпочтениях эксперта суммируются и строится ранжированный ряд;

3. При групповой экспертизе строится обобщенный ранжированный ряд и рассчитываются весовые коэффициенты.

Пример Результат оценки программных продуктов, обозначенных номерами от 1 до 6 представлен в таблице 1.2, где предпочтению i –го продукта над j –м соответствует 1, а противоположному отношению – 0. Требуется расставить продукты по качеству.

Таблица 1.2 - Результат дегустации пищевых продуктов

Продукт	j						Итого
i	1	2	3	4	5	6	
1		1	0	1	1	1	4
2	0		0	1	1	1	3
3	1	1		1	1	1	5
4	0	0	0		0	0	0
5	0	0	0	1		0	1
6	0	0	0	1	1		2

Решение: Ранжированный ряд имеет вид: $Q_3 > Q_1 > Q_2 > Q_6 > Q_5 > Q_4$
или $Q_4 < Q_5 < Q_6 < Q_2 < Q_1 < Q_3$

Вывод: Лучшим по качеству признан продукт №3, худшим – продукт №4.

г) Метод **последовательного сопоставления** предполагает последовательное сравнение каждого объекта экспертизы с совокупностью всех тех, которые ниже рангом. Метод может использоваться совместно с вышеперечисленными, чтобы уточнить позиции входящих ранжированный ряд объектов.

д) **Социологический** метод. Основан на определении мнения различных потребителей оцениваемой продукции (фактических или потенциальных) путем опроса, анкетирования и т.д. (организация конференций, выставок, аукционов, опытно-показательной эксплуатации).

3) **Регистрационный** метод применяется для растянутых по времени процессов. Например, во время испытаний, эксплуатации или сопровождения системы фиксируют определенные события: сбои, отказы, время восстановления после сбоев, длительность работы системы до очередного отказа и т.п. Собранную информацию обрабатывают с применением теоретических и эмпирических зависимостей, методов обработки статистических данных. Метод позволяет определить длительность и точность вычислений, время реакции, необходимые ресурсы и другие характеристики. Для сбора информации и определения показателей качества могут использоваться специальные программные инструменты.

Применение метода **регистрации** событий и/или **расчетов** рассмотрим на примере **оценки надежности**.

Надежность – набор атрибутов, характеризующих способность программного обеспечения сохранять свой уровень качества функционирования в заданных программными документами условиях в течение установленного интервала времени, в том числе и в случае возникновения отклонений в среде функционирования, вызванных сбоями технических средств, ошибками во входных данных, ошибками обслуживания и другими дестабилизирующими воздействиями.

В числе показателей надежности ПП, определяемых путем обработки статистической информации следующие:

– **Вероятность безотказной работы** $P(t_3)$ – показывает вероятность того, что в пределах заданной наработки отказ системы не возникнет;

– **Вероятность отказа** – показатель, обратный предыдущему:

$$Q(t_3) = 1 - P(t_3)$$

где t_3 – заданная наработка, ч.; $Q(t_3)$ – вероятность отказа. Показывает, какова вероятность того, что в пределах заданной наработки отказ системы возникнет.

– **Интенсивность отказов системы** $\lambda(t)$ – число отказов в единицу времени, отнесенное к числу систем (элементов), оставшихся исправными к началу рассматриваемого промежутка времени.

– **Средняя наработка на отказ** T_o – среднее время работы системы (элемента) между двумя соседними отказами. Показывает математическое ожидание времени работы ПП до очередного отказа:

$$T_i = \int_0^t t \cdot f(t) dt$$

где $f(t)$ – плотность вероятности отказа в момент времени t .

$$f(t) = \frac{d}{dt} Q(t) = \frac{d}{dt} (1 - P(t)) = -\frac{d}{dt} P(t)$$

Статистически наработка на отказ определяется отношением суммарной наработки восстанавливаемых объектов к суммарному числу отказов этих объектов:

$$T_o = \frac{\sum_{i=1}^n \tau_i}{n},$$

где n – число отказов системы (элемента) за время испытания; t_i – время безотказной работы системы (элемента) между $(i-1)$ -м и i -м отказами.

– *Среднее время восстановления T_v* – среднее время отыскания и устранения одного отказа. Для определения среднего времени восстановления из статистических данных пользуются формулой

$$T_v = \frac{\sum_{i=1}^n t_i}{n},$$

где n – число отказов за определенный период эксплуатации системы (элемента); t_i – время восстановления системы (элемента) после i -го отказа. Время восстановления складывается из времени, затраченного специалистом по тестированию на выявление отказа и его локализацию; времени устранения отказа; времени последующей проверки работоспособности.

Существуют также другие показатели. Количественные показатели надежности могут использоваться для оценки достигнутого уровня технологии программирования, для выбора метода проектирования будущего программного средства.

2.3. Способы повышения качества

Способы повышения качества ПО можно разделить на проактивные и реактивные.

Проактивные методы направлены на предупреждение возникновения ошибок при создании программных продуктов. К ним относятся:

- повышение квалификации разработчиков;
- внедрение эффективных процессов разработки ПО;
- проектная методология разработки;
- использование стандартов кодирования;
- методологии контроля изменений;
- выравнивание требований к продукту с ожиданиями заказчика.

Реактивные методы применяются к уже созданному ПО и включают инспекции кода, различные методы тестирования, использование статических анализаторов кода, локализацию и исправление найденных ошибок.

Таким образом, тестирование – один из способов повышения качества ПП.

Другая классификация методов, направленных на обеспечение качества программных продуктов: а) экономические (финансирование работ по поддержанию и улучшению качества, поощрение специалистов за создание качественного ПО), б) технические (внедрение современной техники и технологий), в) организационные (например, аудит качества).