5. Kogan, K. Dynamic Generalized Assignment Problems with Stochastic Demands and Multiple Agent-Task Relationships / K. Kogan, E. Khmelnitsky, T. Ibaraki // Journal of Global Optimization, 31, 2005, pp. 17-43.
6. Spivey, M. Z. The Dynamic Assignment Problem / M. Z. Spivey, W. B. Powell // TRANSPORTATION SCIENCE, 38, 4, 2004, pp. 399-419.
7. Ripon, K. S. N. Genetic Algorithm Using a Modified Backward Pass Heuristic for the Dynamic Facility Layout Problem / K. S. N. Ripon, K. Glette, D. Koch, M. Hovin, J. Torresen // PALADYN Journal of Behavioral Robotics, 2 (3), 2011, pp. 164-174.
8. Тихомирова, Д. А. Построение математической модели динамической задачи о назначениях с учетом профессиональных навыков исполнителей / Д. А. Тихомирова, В. В. Мешечкин // Образование, наука, инновации: вклад молодых исследователей – материалы XI (XLIII) Международной научно-практической конференции / Кемеровский государственный университет. – Кемерово: 2016. – Вып. 17. – С. 800-802.
9. Методика оценки уровня квалификации педагогических работников / под ред. В. Д. Шадрикова, И. В. Кузнецовой. – М.: Институт содержания образования государственного университета – Высшей школы экономики, 2010. – 174 с.
10. Данилов, Н. Н. Основы математической теории оптимальных процессов / Н. Н. Данилов, В. В. Мешечкин. – Кемерово: Кузбассвузиздат, 2004. – 219 с.

ОПРЕДЕЛЕНИЕ ОБЪЕМА ВЫЧИСЛИТЕЛЬНЫХ ЭКСПЕРИМЕНТОВ ПРИ РЕШЕНИИ ЗАДАЧ ОПТИМИЗАЦИОННОГО СИНТЕЗА ДИНАМИЧЕСКИХ СИСТЕМ МЕТОДОМ ПЛП-ПОИСКА

И.Н. Статников, к.т.н., с.н.с., Г.И. Фирсов
Институт машиноведения им. А.А. Благонравова РАН, Москва
101990, Москва, Малый Харитоньевский пер., 4, тел. (495) 624-00-72
E-mail: firsovgi@mail.ru

Аннотация: Даются рекомендации по выбору необходимого числа машинных экспериментов при решении конкретных задач исследования и оптимального проектирования методом ПЛП-поиска, позволяющего осуществлять глобальный квазиравномерный просмотр заданной области варьируемых параметров и применить формальные оценки из математической статистики.

Abstract: Recommendations about the choice of necessary number of machine experiments at the solution of specific objectives of a research and optimum design are made by method of the PLP-search allowing to carry out global quasi uniform viewing of the set area of the varied parameters and to apply formal estimates from mathematical statistics.

В методе ПЛП-поиска [1,2], созданном для решения задач оптимального проектирования механизмов и машин на предварительном этапе, важным является статистический анализ результатов, получаемых при моделировании на ЭВМ. Этот анализ целиком базируется на идеях дисперсионного анализа [3], и поэтому его эффективное применение зависит от выбора оценочной функции (критерия значимости) и объема N проводимых на ЭВМ машинных экспериментов. С помощью дисперсионного анализа осуществляют проверку различных статистических гипотез.

В частности, в ПЛП-поиске используется следующая нулевая гипотеза H_0 если средние значения $\tilde{\Phi}_i(\alpha_j)$ критерия качества проектируемой машины статистически не отличаются друг от друга и от общего среднего $\tilde{\Phi}_0(\overline{\alpha})$ всей совокупности N экспериментов, то полагаем, что рассматриваемый исследуемый параметр α_j (j = 1, 2, ..., r) не оказывает в среднем влияния на величину $\Phi(\overline{\alpha})$; при этом i = 1, ... M_j, а M_j - количество уровней, на которые разбивается в ПЛП-поиске диапазон изменения α_j.

Для количественной оценки справедливости выдвигаемой нулевой гипотезы (или ей альтернативной гипотезы H_1) в математической статистике рассматриваются оценочные функции, в том или ином виде используемые при сопоставлении выборочных характеристик, в частности, составляющих дисперсии σ_0^2 всей совокупности проделанных экспериментов (или их несмещенных оценок). Проверка справедливости статистических гипотез при использовании оценочных функций основывается на сравнении некоторого числового показателя, найденного по результатам экспериментов, с табличными значениями оценочных функций при некоторых значениях доверительных вероятностей p, что адекватно доверительному уровню значимости 1 - p. Таблицы теоретических значений вероятностей оценочных функций строятся на основе известных в математической статистике распределений (нормального, биномиального, распределения Рэлея и др.) [3].

Среди множества существующих оценочных функций параметрического типа, применение которых зависит от вида эмпирического распределения наблюденных результатов, в ПЛП-поиске был выбран критерий Фишера F (или дисперсионное отношение), который равен

$$F = s_1^2 / s_2^2, \tag{1}$$

где s - оценка дисперсии. Нужно брать $s_1^2 > s_2^2$. При этом большему значению s_1^2 соответствует число степеней свободы ν_1. Суть формулы (1) состоит в том, что если отношение двух оценок дисперсий s_1^2 и s_2^2 при фиксированных ν_1 и ν_2 и заданном уровне значимости 1 - р больше табличного значения F_T, рассчитанного для тех же условий, то нулевая гипотеза отвергается; если же $F < F_T$, то гипотеза принимается, т. е. полагается, что s_1^2 и s_2^2 несущественно отличаются в статистическом смысле и выборки однородны, т.е. взяты из одной нормальной совокупности (для этих целей рассчитаны таблицы значения критерия F_T, при разных ν_1, ν_2 и р, например, табл. XVIII — XXI из [3]).

Из сказанного ясно, что эффективность применения критерия F зависит от степени приближения распределения результатов экспериментов (вернее, выборочных средних) к нормальному распределению. Распределение средних выборочных носит характер кривой Гаусса, и в этом заложено объективное свойство распределения $\tilde{\Phi}_i(\alpha_j)$. И это свойство не зависит от того, как распределяется сама случайная величина $\Phi(\overline{\alpha})$. Разница состоит лишь в том, что при нормальном распределении самой величины $\Phi(\overline{\alpha})$ распределение средних выборочных $\tilde{\Phi}_i(\alpha_j)$ не зависело бы от объема выборки.

Следует отметить, что результативность использования большинства оценочных функций параметрического типа также существенно зависит от степени близости эмпирического распределения к нормальному. Поэтому на вопрос, какая же из оценочных функций наиболее логически оправдана, в статье Р.К. Бауэра [4] дается следующий ответ: «Этот спор, без сомнения, решается в пользу F - показателя Р.А. Фишера, в котором прямо и попросту сопоставляются обе компоненты общего рассеяния, характеризующие стохастическую модель испытаний». Кроме того, из многих оценочных функций параметрического типа критерий Фишера менее других чувствителен к отклонениям эмпирических распределений от нормального [5].

Естественно, что степень отклонения экспериментальных данных от нормального распределения зависит от объема N экспериментов, более конкретно - от числа выборок. Эта зависимость вытекает из центральной предельной теоремы Ляпунова [6] в теории вероятностей, которая утверждает, что какова бы ни была исходная совокупность, при достаточно большом числе независимых испытаний распределение выборочных средних стремится к нормальному с дисперсией σ_0^2 / N, где σ_0^2 - дисперсия исходной совокупности.

При практическом использовании метода ПЛП-поиска задаются и определяются следующие параметры матрицы планирования экспериментов на ЭВМ: r - число исследуемых параметров α_j, M_j - число уровней; T_i - число реализаций (экспериментов) на i-м уровне j-го параметра. Общее число всех машинных экспериментов при этом равно $N = M_j T_i$, если M_j - const и T_i = const или $N = \sum_{i=1}^{M_j} T_i$, если $T_i \neq$ const.

Число уровней M_j выбирается из следующих соображений. С одной стороны, из соображений равномерного характера распределения [7] M_j должно быть кратно степени двойки, т. е. $M_j = 2^l$ (l = 1, 2,...). С другой стороны, число M_j определяется той тщательностью, с которой исследователь считает необходимым просмотреть заданный диапазон изменения каждого параметра α_j.

В большинстве решавшихся авторами практических задачах проектирования механизмов и исследованиях на тестовых функциях на основе ПЛП-поиска [8] диапазоны варьируемых параметров в основном разбивались на 16 сечений (M_j = 16). Покажем, что в этих задачах распределение средних значений $\tilde{\Phi}_i(\alpha_j)$ также близко к нормальному. Для этого воспользуемся следующим критерием согласия эмпирического распределения с нормальным [9]. Если выборочные асимметрия и эксцесс удовлетворяют неравенствам

$$|A| \le 3\sqrt{D(A)} \text{ и } |E| \le 5\sqrt{D(E)}, \tag{2}$$

то наблюдаемое эмпирическое распределение можно считать нормальным. При этом выборочные асимметрия и эксцесс определяются по следующим формулам:

$$A = \frac{1}{M_j s^3} \sum_{i=1}^{M_j} \left[\tilde{\Phi}_i(\alpha_j) - \tilde{\Phi}_0(\bar{\alpha}) \right]^3, \quad (3)$$

$$E = \frac{1}{M_j s^4} \sum_{i=1}^{M_j} \left[\tilde{\Phi}_i(\alpha_j) - \tilde{\Phi}_0(\bar{\alpha}) \right]^4 - 3. \quad (4)$$

Для расчета дисперсий D(A) и D(E) в [9] приводятся следующие формулы:

$$D(A) = \frac{6(M_j - 1)}{(M_j + 1)(M_j + 3)}, D(E) = \frac{24M_j(M_j - 2)(M_j - 3)}{(M_j + 1)^2 (M_j + 3)(M_j + 5)}, \quad (5)$$

где M_j соответствует объему выборки. Как подчеркивается в [9], приведенный критерий согласия является приближенным, и годится при значениях $M_j \le 20$.

В табл. 1 приведены результаты расчетов по формулам (2) - (5) для практических задач, описанных в [10, 11], и тестовой функции Розенброка $\Phi(\bar{\alpha}) = 100(\alpha_2 - \alpha_1^2)^2 + (1 - \alpha_1)^2; G_0 : \{\alpha_1 \in (-2; 2); \alpha_2 \in (-0,5; 3,5)\}$ [12]. При этом при $M_j = 16$ $3\sqrt{D(A)}$ = 1,58 и $5\sqrt{D(E)}$ =3,89. Известно, что увеличение общей совокупности N машинных экспериментов делает более достоверными те статистические выводы, которые мы стремимся получить.

Такое понимание роли объема N проводимых на ЭВМ экспериментов соответствует и интуитивному желанию при использовании ПЛП-поиска получить определенные статистические оценки при достаточно полном дискретном обзоре пространства исследуемых параметров, в отличие от ситуации, когда исследователь a priori знает или подозревает область нахождения оптимальных решений. Поэтому для определения общего числа N проводимых на ЭВМ экспериментов нужно привлечь следующие рассуждения. С одной стороны, при планировании экспериментов нижняя граница количества опытов, проводимых в r-м пространстве параметров, должна удовлетворять следующему соотношению [5]:

$$N^* \ge r + C_r^k, \quad (6)$$

где C_r^k - число сочетаний (взаимодействий) из r элементов по k (k = 2, ..., r - 1). С другой стороны, при процессе производства экспериментов на ЭВМ с использованием $ЛП_\pi$-сеток можно полагать, что дисперсия воспроизводимости машинных экспериментов равна нулю (при условии, что ЭВМ работают в режимах, предусмотренных техническими условиями).

Поэтому, учитывая, что в ПЛП-поиске процесс дискретного обзора пространства исследуемых параметров рандомизирован, можно говорить о вероятности P нахождения лучших решений в области, составляющей L-ю часть исходного пространства после проведения N** экспериментов. Три указанные величины связываются следующим уравнением [12]:

$$P = 1 - (1 - L)^{N^{**}}. \quad (7)$$

Учитывая, что для задач проектирования механизмов и машин чаще всего имеет смысл рассматривать лишь эффекты взаимодействий параметров первого и, реже, второго порядка (k = 2, 3), получаем по формуле (6) для $r \le 40$ и k = 2, что $N^* \ge 820$, а для $r \le 17$ и k = 2, 3 - $N^* \ge 832$. В то же время наиболее интересная в практическом отношении часть табл. 2 (выделена жирной линией) для различных P и L содержит значения N** от 22 до 850, которые можно рассматривать как верхнюю границу необходимого числа экспериментов. При этом, конечно, остаются в силе и соображения стоимостного характера. Например, при r = 10 и k = 2 $N^* = 55$. Из табл. 2 видно, что при N**= 70 мы с вероятностью $P \ge 0,98$ будем уверены, что область, содержащая лучшие решения, составит не более 5% от исходной области поиска. Если же мы хотим, чтобы $L \le 0,01$ от исходной области при той же вероятности $P \ge 0,98$, то N** = 425; при этом, если данное количество экспериментов нас не устраивает по стоимостным и временным причинам (а часто первая причина — следствие второй), то помня, что $N^* = 55$, можно назначить N** = 175.

Однако теперь вероятность того, что $L \le 0,01$ от исходной области поиска ниже, чем в предыдущем случае, но все еще практически приемлема ($P \ge 0,80$). Слово «практически» употреблено в том смысле что вся процедура метода ПЛП-поиска рекомендуется к использованию на предварительном этапе решения задачи проектирования технического устройства (научно-исследовательском), когда основная цель состоит не только в достижении абсолютных результатов.(что не отвергается), а в получении объективной информации о свойствах исследуемых параметров по отношению к критериям качества проекта. Отметим также, что, выбрав число N из диапазона N* ≤ N ≤ N**, руководствуясь всеми изложенными выше соображениями, мы тем самым однозначно определяем число перестановок T_i, необходимое для образования матрицы планирования экспериментов: $T_i = N/ M_j$.

Таблица 1

Оценки выборочной асимметрии и эксцесса для параметров математических моделей

Параметры α_j		α_1	α_2	α_3	α_4	α_5	α_6	α_8	α_{10}	α_{11}
Пневмовстряхивающая машина	$\lvert A\rvert$	0,19	0,74	0,18	0,44	0,07		0,28	0,35	0,05
	$\lvert E\rvert$	1,59	1,01	2,20	1,27	1,38		0,80	1,30	1,28
Зубчато-рычажный механизм	$\lvert A\rvert$	0,62	0,32	0,83	0,82	0,05	0,55			
	$\lvert E\rvert$	0,28	0,82	0,03	0,38	1,18	0,12			
	$\lvert A\rvert$	0,58	0,08	0,58	0,02	0,10	0,44			
	$\lvert E\rvert$	0,30	0,96	1,06	1,76	1,46	1,46			
	$\lvert A\rvert$	0,94	0,68	0,83	0,07	0,07	0,14			
	$\lvert E\rvert$	0,52	0,77	0,46	1,52	0,76	1,07			
Функция Розенброка	$\lvert A\rvert$	1,19	0,95							
	$\lvert E\rvert$	0,51	0,95							

Некоторые результаты расчетов по уравнению (7) приведены в табл. 2.

Исходя из рекомендаций математической статистики [3], следует стремиться к тому, чтобы $T_i \geq 10$, что также может явиться дополнительным аргументом при выборе числа N. При r > 40 и k = 2 и при r > 17 для k = 2, 3 число N*, определяемое по формуле (6), достаточно близко по величине к крайним значениям N** в табл. 2, поэтому в таких задачах это число и следует принимать в качестве N = N*. Таким образом, вопрос об априорном назначении объема N вычислительных экспериментов не может быть разрешен для всех задач вообще, так как значение N в первую очередь зависит от вида поверхности $\tilde{\Phi}_i(\alpha_j)$, что чаще всего неизвестно перед началом экспериментов. Однако несколько общих моментов в подходе к выбору значения можно высказать.

Таблица 2

Оценка числа экспериментов N**

P	L							
	0,20	0,15	0,10	0,05	0,03	0,01	0,005	0,001
0,70	5	7	12	22	40	130	260	1300
0,75	6	8	13	25	46	150	300	1500
0,80	7	10	16	29	54	175	350	1750
0,85	9	12	18	34	63	205	410	2050
0,90	10	14	22	41	77	250	500	2500
0,95	14	18	30	54	100	325	650	3250
0,98	18	26	38	70	130	425	850	4250
0,99	21	29	44	83	154	500	1000	5000
0,995	24	34	51	96	177	575	1150	5750

1) Исходя из возможности описать стационарную область концентрации наилучших решений $G_k(\bar{\alpha})$ некоторым полиномом степени d (регрессионная зависимость), нижнюю границу N^* можно определить следующим образом [5]:

$$N^* \geq C_{r+d}^{d}. \qquad (8)$$

2) На основе априорных физических представлений о чувствительности k-го критерия качества к вариациям j-го параметра можно потребовать, чтобы

$$\delta\alpha_j \approx \Delta\alpha_j / M_j. \quad (9)$$

Естественно ориентироваться на максимальное значение M_j, полученное по этой формуле.

3) Пространство в исходной области $G(\bar{\alpha})$ изменения параметров, объем которой $\delta V = \prod_1^r \delta\alpha_j$, будем рассматривать как элементарную r-мерную ячейку. Тогда, исходя из свойств квазиравномерного распределения точек в области $G(\bar{\alpha})$ [5], можно считать, что $(\delta V / V) \approx N_0^{'} / N$, где V- объем области $G(\bar{\alpha})$, $N_0^{'}$ - число точек из N, попавших в элементарную ячейку. Справедливым будет при квазиравномерном распределении точек $\bar{\alpha}$ в области $G(\bar{\alpha})$ рассматривать значение U = $N_0^{'} / N$ как вероятность обнаружения экстремального значения k-го критерия качества Φ_k^+ с одного испытания (эксперимента). Очевидно, что с ростом r значение U достаточно мало. Тогда высокие значения вероятности P обнаружения $\Phi_k(\bar{\alpha}) = \Phi_k^+$ после проведения N^{**} экспериментов определяются соотношением

$$P = 1 - (1 - U)^{N^{**}}. \quad (10)$$

Естественно полагать, что стационарная область, для которой справедливо ее описание полиномом степени d (а по существо это есть разложение $\Phi_k(\bar{\alpha})$ в точке $\bar{\alpha}^+$ в ряд Тейлора), никак не меньше области, рассматриваемой как элементарная ячейка. Отсюда можно считать, что общее число точек N определяется двумя границами

$$N^* \le N \le N^{**}, \quad (11)$$

где N^* и N^{**} определяются соответственно из (8) и (10).

4) Следует заметить, что на выбор N^{**}, а значит, и N может повлиять и то обстоятельство, что в многокритериальной задаче значения критериев качества могут быть сильно коррелированы между собой. Тогда вероятность P из (10) должна быть заменены условной вероятностью типа $P(\Phi_k^* | \Phi_{k-1}^*)$ и т.д. На выбор N из (11), конечно, влияют временные и стоимостные аспекты проведения данных вычислительных экспериментов.

5) По поводу назначения самой величины M_j, из (9) следует подчеркнуть, что там, где это возможно, следует брать M_j, кратный степени двойки, что способствует квазиравномерности распределения точек $\bar{\alpha}$ [7].

Надежным основанием при выборе N с учетом рассмотренных обстоятельств должен служить небольшой по объему предварительный эксперимент на ЭВМ, в котором на основе информационного подхода будут получены энтропийные оценки области $G(\bar{\alpha})$ по каждому критерию $\Phi_k(\bar{\alpha})$ [13, 14].

Литература

1. Статников И.Н., Фирсов Г.И. Интеллектуальный метод решения задач моделирования и исследования машин и механизмов на основе планируемого вычислительного эксперимента // Актуальные проблемы современного машиностроения. Международная научно-практическая конференция. (Юрга,11-12 декабря 2014 г.). Сборник трудов. – Томск: Изд-во Томского политехнического университета, 2014. – С. 157-161.
2. Статников И.Н., Фирсов Г.И. Использование интеллектуальных методов решения задач проектирования современных машин и механизмов // Актуальные проблемы современного машиностроения. Международная научно-практическая конференция. (Юрга, 17-18 декабря 2015 г.). Сборник трудов. – Томск: Изд-во Томского политехнического университета, 2015. – С. 63-68.
3. Митропольский А.К. Техника статистических вычислений. М.: Наука, 1971. - 576 с.
4. О теории дисперсий. Сборник статей. Составитель Н.С. Четвериков. - М.: Статистика, 1968. - 239 с.
5. Налимов В.В. Теория эксперимента. - М.: Наука, 1971. - 208 с.
6. Дружинин Н.К. Логика оценки статистических гипотез. - М.: Статистика, 1973. - 212 с.
7. Соболь И.М. Многомерные квадратурные формулы и функции Хаара. - М.: Наука, 1969. - 288 с.
8. Статников И.Н.. Андреенков Е.В. ПЛП-поиск – эвристический метод решения задач математического программирования. – М.: ИИЦ МГУДТ, 2006. – 140 с.
9. Пустыльник Е.И. Статистические методы анализа и обработки наблюдений. - М.: Наука, 1968. - 288 с.

10. Статников И.Н., Фирсов Г.И. Численный подход к решению задачи рационального выбора параметров динамических систем // Вестник научно-технического развития. – 2016. - № 3(103). - С.38-53.
11. Статников И.Н., Фирсов Г.И. Интерактивное структурирование пространства параметров при проектировании динамических систем // Вестник Тамбовского государственного технического университета. – 2015. - № 1. – С.36-41.
12. Уайлд Д.Дж. Методы поиска экстремума. - М.: Наука, 1967. - 268 с.
13. Статников И.Н., Фирсов Г.И. Методы энтропийного анализа в задачах оптимального проектирования методом ПЛП-поиска // Современные проблемы информатизации в моделировании и программировании. Сборник трудов. Вып 11. (по итогам XI международной открытой научной конференции). - Воронеж: Научная книга, 2006. - С.205-206.
14. Статников И. Н., Фирсов Г. И. Процедура структурирования пространства параметров динамической системы по энтропийному критерию при анализе многокритериальных задач проектирования // Информационные технологии и математическое моделирование (ИТММ-2016). XV Международная научно-практическая конференция. (Анжеро-Судженск, 12-16 сентября 2016 г.). Материалы. Часть 2. - Томск: Изд-во Томского ун-та, 2016. - С.215-220.

ПРОБЛЕМА ТРУДОУСТРОЙСТВА МОЛОДЫХ СПЕЦИАЛИСТОВ НА РЫНКЕ ТРУДА (НА ПРИМЕРЕ КЕМЕРОВСКОЙ ОБЛАСТИ)

И.В .Добрычева, ассистент кафедры ЭиАСУ
И.С Богданов А.В. Большанин, студенты группы 17б51,
Юргинский технологический институт (филиал) Томского политехнического университета
652055, Кемеровская обл., г. Юрга, ул. Ленинградская, 26

Аннотация: В данной статье рассматривается проблема трудоустройства молодых специалистов в условиях современных тенденций развития рыночной экономики. Безработица является одной из самых важных проблем в сфере труда. Целью статьи является анализ и выявление проблем молодежной безработицы.

Сегодня, в условиях мирового кризиса, молодым специалистам очень тяжело найти работу с нормальными условиями труда, хорошей заработной платой, а также соответствующую их специальности. С данной проблемой сталкивается в наши дни огромное количество человек, находящихся в поиске работы. На сегодняшний день численность безработных граждан, состоящих на учете в Кемеровском Центре Занятости Населения на 16,9% больше, чем на аналогичную дату предыдущего года.[1]

Рассмотрим состав занятого населения Кемеровской области по уровню образования за 2014 и 2015 гг. (Рисунки 1, 2).

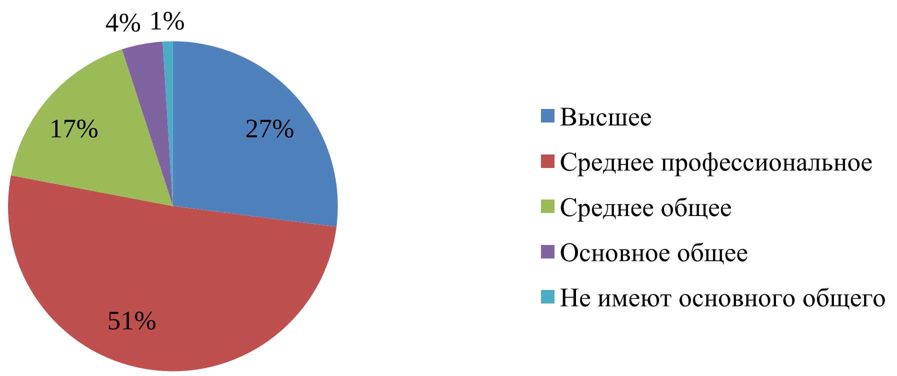

Рис. 1. Состав занятого населения по уровню образования в Кемеровской области в 2014 г.