МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего профессионального образования
Набережночелнинский институт (филиал)
"Казанский (Приволжский) федеральный университет"

АНАЛИЗ ДАННЫХ ЭКСПЕРИМЕНТА
И ОПТИМИЗАЦИЯ В ТЕХНИЧЕСКИХ ЗАДАЧАХ

Методические указания к практическим занятиям
по курсу “Основы научных исследований”

Набережные Челны
2018 г

УДК 629.113

Печатается по рекомендации к изданию, принятой на заседании кафедры «Автомобили, автомобильные двигатели и дизайн» Набережночелнинского института (филиала) КФУ от "_15_" _февраля_ 2018 г., протокол № 13.

Рецензент д-р техн. наук, проф. Кулаков А.Т.

Анализ данных эксперимента и оптимизация в технических задачах. Метод. указания к практическим занятиям / Сост.: **А.Ю. Барыкин, Р.Р. Басыров, М.М. Мухаметдинов. -** Наб. Челны: НЧИ КФУ, 2018. - 38 с.

Методические указания посвящены прикладным вопросам анализа результатов экспериментальных и теоретических исследований в технических дисциплинах. Приведены расчётные методики и примеры исследований автотранспортных средств.

Методические указания могут быть использованы в учебном процессе при изучении дисциплине "Основы научных исследований" и в дипломном проектировании для студентов вузов, обучающихся по направлениям подготовки: 23.04.02 «Наземные транспортно-технологические комплексы» (квалификация выпускника: магистр), 23.05.01 «Наземные транспортно-технологические средства» (квалификация выпускника: инженер), 13.04.03 «Энергетическое машиностроение» (квалификация выпускника: магистр).

Данные методические указания также могут быть использованы в учебном процессе по аналогичной дисциплине направлений подготовки 23.03.03 «Эксплуатация транспортно-технологических машин и комплексов» (квалификация выпускника: бакалавр), 23.03.01 «Технология транспортных процессов» (квалификация выпускника: бакалавр).

Табл. 13.
Рис. 1.
Библиогр. 15 назв.

В соответствии с программой в результате изучения дисциплины “Основы научных исследований” студент, обучающийся по направлениям подготовки: 23.04.02 «Наземные транспортно-технологические комплексы» (квалификация выпускника: магистр), 23.05.01 «Наземные транспортно-технологические средства» (квалификация выпускника: инженер), 13.04.03 «Энергетическое машиностроение» (квалификация выпускника: магистр), должен уметь получить, обработать и анализировать результаты исследований.

Отечественными и зарубежными учёными разработаны многие десятки математических моделей колёсных машин, отличающиеся различной степенью соответствия реальным физическим процессам. Высокая сложность и многообразие явлений, возникающих при движении автомобиля, в большинстве случаев не позволяют провести сугубо теоретическое описание. Прикладной характер исследований подразумевает достаточно широкое использование экспериментальных данных и эмпирических зависимостей.

Необходимо также учитывать, что входные возмущения тех или иных факторов разнообразны по своим характеристикам и могут иметь случайный или последовательно изменяющийся характер. В некоторых случаях известны только предельные значения параметров из-за недостаточной исследованности процесса - по причине крайней сложности или из-за недостатка информации.

1. Анализ результатов эксперимента. Обоснование количества опытов.

<u>Методика планирования эксперимента</u> - важнейшая составная часть плана эксперимента. В ней излагаются: последовательность действий; основные приёмы и правила осуществления каждого этапа, использование приборов и оборудования; порядок измерения, фиксации результатов и методы их обработки; порядок анализа результатов эксперимента.

При разработке методики важно правильно обосновать количество опытов, которое гарантирует требуемую точность результата, а с другой стороны – не ведёт к неоправданному перерасходу средств и времени на избыточные испытания.

Анализ случайных погрешностей основывается на теории случайных ошибок. Эта теория даёт возможность с определённой гарантией оценить возможные ошибки и вычислить истинное значение измеряемой величины. В основе теории случайных ошибок лежит предположение о том, что при бесконечно большом числе измерений истинное значение измеряемой величины равно среднеарифметическому значению всех результатов измерений, а случайные погрешности одинаковой величины, но разного знака встречаются одинаково часто. При этом появление того или иного результата измерения как случайного события описывается нормальным законом распределения.

Для нормального закона распределения общей оценочной характеристикой измерений является дисперсия:

$$D_0 = \sigma_0^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1};$$

где σ_0 - среднее квадратичное отклонение измеренных значений x_i от среднеарифметического $\overline{x}$; n - число измерений.

Дисперсия характеризует однородность измерения: чем выше D_0, тем больше разброс измерений.

Доверительным интервалом называется интервал значений x_i, в который с заданной вероятностью попадает истинное значение x_∂ измеряемой величины. Вероятность того, что истинное значение x_∂ находится в данном доверительном интервале, называется доверительной вероятностью измерения P_∂. Наиболее часто доверительная вероятность P_∂ принимается равной $0{,}90$; $0{,}95$; $0{,}99$. Доверительный интервал характеризует точность измерений данной серии, а доверительная вероятность - достоверность измерений.

При небольшом числе измерений ($n < 20$) доверительный интервал μ рассчитывается из выражения:

$$\mu = t \cdot \sigma_0 ;$$

где t - значение критерия Стьюдента, выбираемое в зависимости от принятой доверительной вероятности P_∂ и числа измерений (табл. 3); σ_0 - среднеарифметическое значение среднеквадратичного отклонения σ, определяется по формуле:

$$\sigma_0 = \sigma / \sqrt{n} .$$

Истинное значение измеряемой величины имеет вид:

$$x_\partial = \bar{x} \pm \mu .$$

Относительная погрешность серии измерений при заданной доверительной вероятности определяется по формуле:

$$\delta = \frac{\mu}{\bar{x}} \cdot 100\% .$$

Если расчётная величина доверительного интервала соизмерима с погрешностью определения δ и $\mu \le 3 \cdot \delta$, то скорректированное значение доверительного интервала определяется из выражения:

$$\mu' = \sqrt{\mu^2 \pm \left(\frac{t_\infty \cdot \delta}{3} \right)^2} ;$$

где t_∞ - табличное значение критерия Стьюдента при заданной доверительной вероятности P_∂ и бесконечно большом числе измерений $n = \infty$ (табл. 1).

При количестве испытаний более десяти <u>методика обоснования количества опытов</u> базируется на неравенстве Чебышева:

$$P[| \bar{x} - M(x) | \le \varepsilon] \le 1 - \frac{D_0(x)}{n \cdot \varepsilon^2} ; \qquad (1)$$

“Вероятность того, что разница между среднестатистическим $\bar{x}$ и математическим ожиданием $M(x)$ не превысит ε, равна разности между единицей и отношением $D_0(x) / n \cdot \varepsilon^2$ ”;

где n - количество проведённых опытов; $\bar{x}$ - среднее значение случайно измеряемой в ходе эксперимента величины x; $M(x)$ - математическое ожидание величины x ($M(x)=\lim\limits_{n\to\infty}\bar{x}$); $D_0(x)$ - дисперсия величины x, рассчитанная по результатам n опытов; ε - требуемая (заданная) точность результата.

Табл. 1.
Критерий Стьюдента.

n	t при P_∂		
	0,90	0,95	0,99
2	6,31	12,71	63,70
3	2,92	4,30	9,92
4	2,35	3,18	5,84
5	2,13	2,77	4,60
6	2,02	2,57	4,03
7	1,94	2,45	3,71
8	1,90	2,36	3,50
9	1,86	2,31	3,36
10	1,83	2,26	3,25
15	1,76	2,14	2,98
20	1,73	2,09	2,86
∞	1,64	1,96	2,58

В этой формуле имеется три неизвестных: n и статистические характеристики $\bar{x}$ и $D_0(x)$, зависящие от n. Поэтому процесс расчёта n является итеративным: вначале задаётся некоторое априорное значение n, проводится n-ное количество опытов, вычисляется $D_0(x)$ и проверяется неравенство (1). Если оно выполняется, то количество опытов достаточно. В противном случае количество опытов увеличивается.

Пример. *Экспериментальные данные испытаний шипа крестовины межколёсного дифференциала на смятие (предельное напряжение смятия $\sigma_{см}$, МПа) приведены в табл. 2.*

Статистические характеристики рассчитываются по известным формулам математической статистики:

Табл. 2.
Опытные данные.

№ опыта (i)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$\sigma_{см}$	12,42	12,69	12,81	12,63	12,56	12,19	12,78	12,45	12,40	12,22

$$\bar{x} = 12{,}515\,;$$

$$D_0(x) = \sum_{i=1}^{n} (x_i - \bar{x})^2 \Big/ n = 0{,}0422\,.$$

По результатам испытаний определяется их требуемое число, которое обеспечит с заданной вероятностью 0,95 разницу между $\bar{x}$ и $M(x)$ не более 0,1 МПа (т.е. $\varepsilon = 0{,}1$). Из неравенства Чебышева найдём искомую величину:

$$0{,}9 \le 1 - \frac{0{,}0422}{n \cdot 0{,}2^2}\,;$$

$$n \ge \frac{0{,}0422}{0{,}1^2 \cdot (1 - 0{,}9)} = 42{,}2\,.$$

Требуемое количество испытаний превышает 42 опыта. Проведя дополнительные 33 опыта, нужно снова рассчитать $\bar{x}$ и $D_0(x)$ и проверить требуемый показатель n. Если неравенство Чебышева будет соблюдено, то испытания можно прекратить. В противном случае проводятся дополнительные испытания, пока неравенство (1) не будет выполнено.

(Можно при фиксированном числе испытаний определить получаемую разницу между $\bar{x}$ и $M(x)$. В данном случае (при 10 опытах) она составляет 0,205 МПа.)

2. Исключение грубых ошибок. Сглаживание данных эксперимента.

При анализе результатов эксперимента следует исключить грубые ошибки (промахи). Однако, прежде чем исключить то или иное измерение, необходимо убедиться, что это действительно грубая ошибка. Наличие промахов определяется по критериям:

$$\beta_1 = \frac{x_{max} - \bar{x}}{\sigma \cdot \sqrt{(n-1)/n}}; \quad \beta_2 = \frac{\bar{x} - x_{min}}{\sigma \cdot \sqrt{(n-1)/n}};$$

где x_{max}, x_{min} - наибольшее и наименьшее значения из n измерений.

Максимальное значение критерия β_{max}, возникающее вследствие статистического разброса, определяется по табл. 3 в зависимости от числа измерений n и заданной доверительной вероятности P_∂.

Если $\beta_1 > \beta_{max}$, то значение x_{max} следует исключить из серии измерений как промах. При $\beta_2 < \beta_{max}$ исключается значение x_{min}. После исключения грубых ошибок определяют новые значения $\bar{x}$ и σ из $n-1$ или $n-2$ измерений.

Табл. 3.

Критерий β_{max}.

n	β_{max} при P_∂		
	0,90	0,95	0,99
3	1,41	1,41	1,41
4	1,64	1,69	1,72
5	1,79	1,87	1,96
6	1,89	2,00	2,13
7	1,97	2,09	2,26

8	2,04	2,17	2,37
9	2,10	2,24	2,46
n	β_{max} при $P_д$		
	0,90	0,95	0,99
10	2,15	2,29	2,54
15	2,33	2,49	2,80
20	2,45	2,62	2,96

Пример. Определить истинное значение максимальной скорости автомобиля ВАЗ-2106 и относительную погрешность серии из 10 измерений при доверительной вероятности 0,95 и погрешности определения $\pm 0{,}1\%$. Результаты измерений приведены в табл. 4.

Табл. 4.
Данные измерений.

№ опыта	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
v_A, **км/час**	148,1	148,3	149,4	147,4	149,0	150,1	148,5	148,2	147,9	149,3

Расчёт исходных величин приведен в табл. 5 и ниже.

Табл. 5.
Расчётные значения.

№ определения	x_i	$\bar{x} - x_i$	$(\bar{x} - x_i)^2$
1	*148,1*	*0,52*	*0,2704*
2	*148,3*	*0,32*	*0,1024*
3	*149,4*	*-0,78*	*0,6084*
4	*147,4*	*1,22*	*1,4884*
5	*149,0*	*-0,38*	*0,1444*
6	*150,1*	*-1,48*	*2,1904*
7	*148,5*	*0,12*	*0,0144*
8	*148,2*	*0,42*	*0,1764*
9	*147,9*	*0,72*	*0,5184*
10	*149,3*	*-0,68*	*0,4624*

	1486,2	0,0	5,976

$$\bar{x} = \sum x_i / n = 148{,}62; \quad \sigma = \sqrt{\frac{5{,}976}{10-1}} = 0{,}8149.$$

Анализируем серию на наличие грубых ошибок:

$x_{max} = 150{,}1; \quad x_{min} = 147{,}4;$

$$\beta_1 = \frac{150{,}1 - 148{,}62}{0{,}8149 \cdot \sqrt{10-1/10}} = 1{,}914$$

$$\beta_2 = \frac{148{,}62 - 147{,}4}{0{,}8149 \cdot \sqrt{10-1/10}} = 1{,}578$$

Для $P_д = 0{,}95$ *и* $n = 10$ *находим в табл. 3:* $\beta_{max} = 2{,}29$. *Исключаем как грубую ошибку* $x_4 = 147{,}4$.

Определяем новые значения $\bar{x}$ *и* σ *из оставшихся 9 измерений (табл. 6).*

Табл. 6.
Статистические выкладки.

№ определения	x_i	$\bar{x} - x_i$	$(\bar{x} - x_i)^2$
1	148,1	0,656	0,4303
2	148,3	0,456	0,2079
3	149,4	-0,644	0,4147
4	149,0	-0,244	0,0595
5	150,1	-1,344	1,8063
6	148,5	0,256	0,0655
7	148,2	0,556	0,3091
8	147,9	0,856	0,7327
9	149,3	-0,544	0,2959
Σ	1338,8	0,004	4,3219

$$\bar{x} = \sum x_i / n = 148{,}756;$$

$$\sigma = \sqrt{\frac{4{,}3219}{9-1}} = 0{,}7350\,.$$

Для $P_д = 0{,}95$ *и* $n = 9$ *находим (табл. 1)* $t = 2{,}31$.

Определяем величину доверительного интервала:

$$\mu = 0{,}7350 \cdot 2{,}31 / \sqrt{9} = 0{,}566\,.$$

Величина доверительного интервала превышает погрешность определения.

Тогда истинное значение измеряемой величины:

$v_A = 148{,}756 \pm 0{,}566$ *км/час.*

Относительная погрешность серии измерений:

$$\delta = \frac{0{,}566}{148{,}756} \cdot 100\% = 0{,}38\%\,.$$

<u>Сглаживание данных эксперимента</u> является специальной операцией усреднения с помощью интерполяционных полиномов, обеспечивающей получение уточнённого значения $\bar{y}_i$ по заданному значению y_i и ряду близлежащих значений (..., y_{i-1}, y_i, y_{i+1} ,...), известных со случайной погрешностью.

Линейное сглаживание по трём точкам реализуется с помощью следующих формул:

$$\bar{y}_0 = (5 \cdot y_0 + 2 \cdot y_1 - y_2)/6\,;$$

$$\bar{y}_i = (y_{i-1} + y_i + y_{i+1})/3\,, \quad 1 \le i \le n-1\,;$$

$$\bar{y}_n = (5 \cdot y_n + 2 \cdot y_{n-1} - y_{n-2})/6\,;$$

где n - номер последней точки (ординаты y_i).

Линейное сглаживание по 5 точкам проводится с использованием формул:

$$\bar{y}_0 = (3 \cdot y_0 + 2 \cdot y_1 + y_2 - y_4)/5\,;$$

$$\bar{y}_1 = (4 \cdot y_0 + 3 \cdot y_1 + 2 \cdot y_2 + y_3)/10\,;$$

$$\bar{y}_i = (y_{i-2} + y_{i-1} + y_i + y_{i+1} + y_{i+2})/5\,, \quad 2 \le i \le n-2\,;$$

$$\bar{y}_{n-1} = (y_{n-3} + 2 \cdot y_{n-2} + 3 \cdot y_{n-1} + 4 \cdot y_n)/10\,;$$

$$\bar{y}_n = (3 \cdot y_n + 2 \cdot y_{n-1} + y_{n-2} - y_{n-4})/5 .$$

Пример. Плотность рабочей жидкости вязкостной муфты находится в линейной зависимости от температуры. Результаты сглаживания грубых данных эксперимента (y_i) по методам трёх и пяти точек ($\bar{y}_i$), а также точные значения функции даны ниже в табл. 7.

Табл. 7.
Результаты линейного сглаживания.

i	y_i	$\bar{y}_i$ для сглаживания по трём точкам	$\bar{y}_i$ для сглаживания по пяти точкам	Точное значение $y = 977 - 0{,}861 \cdot x$, кг/м^3, для диапазона x 0…180 ^{0}C
0	1023	989,83	997	977,0
1	902	968,33	965,2	959,78
2	980	925,33	933,4	942,56
3	894	914	914	925,34
4	868	896	906,6	908,12
5	926	886,33	886,8	890,9
6	865	890,67	883,4	873,68
7	881	874,33	876,6	856,46
8	877	864	859,4	839,24
9	834	840,5	842,2	822,02
$\bar{\delta}$,%	3,619	1,617	1,356	-

3. Построение гистограмм.

Одномерный массив из n некоторых цифровых данных x_i характеризуется совокупностью статистических характеристик (одномерная статистика), перечисленных ниже.

Начальные моменты k-того порядка:

$$m_k(x) = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^k \ .$$

Обычно $k = 1{,}2{,}3{,}4$ (точность вычисления m_k при $k > 4$ низкая).

Центральные моменты *k*-того порядка:

$$M_k(x) = \frac{1}{n} \cdot \sum_{i=1}^{n} [x_i - m_i(x)]^k \ ;$$

Момент $M_1(x) = 0$.

Связь центральных моментов с начальными устанавливается соотношениями (аргумент x в скобках опускаем):

$$M_2 = m_2 - m_1^2 \ ;$$

$$M_3 = m_3 - 3 \cdot m_1 \cdot m_2 + 2 \cdot m_1^3 \ ;$$

$$M_4 = m_4 - 4 \cdot m_1 \cdot m_3 + 6 \cdot m_1^2 \cdot m_2 - 3 \cdot m_1^4 \ ;$$

которые позволяют вычислять M_k по мере ввода x_i (без запоминания массива x_i).

Среднее значение:

$$\bar{x} = m_1 = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \ ;$$

есть наиболее вероятное значение числа в массиве.

Дисперсия смещённая:

$$D = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 \ ;$$

является наиболее вероятной степенью отклонения x_i от среднего значения $\bar{x}$.

Стандартное смещённое отклонение $\sigma = \sqrt{D}$ определяет среднеквадратичную погрешность x_i, если за точное значение принять $\bar{x}$.

Дисперсия несмещённая:

$$D_0 = \frac{M_2 \cdot n}{n - 1} = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 ;$$

применяется при статистической обработке чисел x_i с нормальным распределением.

Стандартное несмещённое отклонение (среднее квадратичное отклонение) $\sigma_0 = \sqrt{D_0}$.

Коэффициент ассимметрии:

$$A = \frac{1}{n \cdot D_0^{3/2}} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^3 = \frac{M_3}{M_2^{3/2}} ;$$

характеризует скошенность графической функции плотности распределения вероятностей $P(x)$. При $A = 0$ она симметрична, при $A > 0$ вытянут правый, а при $A < 0$ - левый участок спада кривой $P(x)$.

Коэффициент эксцесса:

$$E = \frac{1}{n \cdot D_0^2} \cdot \sum_{i=1}^{n} (x_i - \bar{x}) = \frac{M_4}{M_2^2} - 3 ;$$

характеризует степень остроты пика кривой $P(x)$ в сравнении с $P(x)$ для нормального распределения. Если $E > 0$, $P(x)$ имеет более острый пик, если $E < 0$ - пик менее острый.

Вспомогательные коэффициенты:

$$\alpha_3 = U_3 = \sqrt{\frac{6 \cdot (n-1)}{(n+1) \cdot (n+3)}} ;$$

$$\alpha_4 = U_4 = \sqrt{\frac{24 \cdot (n-2) \cdot (n-3)}{(n-1)^2 \cdot (n+3) \cdot (n+5)}} ;$$

служат для приближённой проверки гипотезы о нормальном распределении x_i . Если:

$$A < \alpha_3 / (2...3) \text{ и } E < \alpha_4 / (2...3) ; \qquad (2)$$

то распределение $P(x)$ для массива x_i можно считать нормальным.

При программировании вычислений одномерной статистики желательно предусмотреть следующие возможности:

- накопление сумм x_i^k;
- возможность исключения ошибочно введённого числа x_i;
- подсчёт n в ходе ввода x_i;
- выдачу статистических характеристик в любой момент (до окончания ввода всех x_i).

<u>Среднее геометрическое</u> для n чисел x_i определяется как:

$$G_M = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} .$$

<u>Гармоническое среднее:</u>

$$H_M = n \big/ (x_1^{-1} + x_2^{-1} + \ldots + x_n^{-1}).$$

Гистограмма распределения характеризует количество чисел x_i, попадающих в интервалы изменения x с границам d_0, d_1, d_2, … , d_n. Гистограмма распределения графически строится в виде столбцов, высота которых соответствует количеству x_i, попавших в интервал изменения x, на который опирается столбик (на горизонтальной оси x).

Гистограмма интегрального распределения характеризует количество чисел, попадающих в интервалы $(-\infty, d_1)$, $(-\infty, d_2)$, … , $(-\infty, d_n)$.

Простейший алгоритм подготовки гистограмм заключается в сравнении x_i с сеткой границ d_0, d_1, d_2, … , d_n с помощью операций условных переходов и подсчёте числа попаданий в каждый интервал. Однако программа при большом числе интервалов получается громоздкой, увеличивается время обработки каждого x_i. В большинстве случаев отрезки $[d_1, d_2]$, $[d_2, d_3]$ и т.д. имеют одинаковую ширину:

$$C = \Delta x = (x_M - x_0) / n ;$$

где x_M - максимальное значение x_i, x_0 - минимальное значение x_i, n - число интервалов.

Тогда алгоритм ускоренной подготовки данных для построения гистограмм будет следующий.

1. Вводим x_M, x_0 и n.

2. Определяем массив счётчиков $A(n)$, вычисляем C по формуле и полагаем $i = 1$.

3. Организуем ввод x_i, вычисление $i \leftarrow i + 1$ (формирование текущего номера i каждого x_i) и вычисление вспомогательной переменной:
$Y = int(C \cdot (x - x_0))$.

Значение Y указывает на номер того отрезка, в который попадает заданное значение x_i. Функция $int(x)$ вычисляет целую часть числа x.

4. Вычисляем $A(Y) \leftarrow A(Y) + 1$ (т.е. вносим единицу в содержимое счётчика $A(Y)$) и возвращаемся к п. 3.

5. Вывод данных для построения гистограмм организуем с помощью цикла с управляющей переменной j, меняющейся от значения 1 до n с шагом 1. Числа попаданий x_i в j-тый отрезок получаем выводом значений $A(j)$. Для получения данных для построения гистограммы интегрального распределения вычисляем $S \leftarrow S + A(j)$ при начальном $S = 0$.

Статистические расчёты на специализированных калькуляторах обычно выполняются микропрограммно.

Поскольку экспериментальные данные, как правило, носят вероятностный характер, прежде всего, необходимо определить закон распределения исследуемой величины и его характеристики. Для определения закона распределения случайной величины x_i можно воспользоваться широко известными методами статистической проверки гипотез с помощью критериев Пирсона (χ^2), Фишера (F), Колмогорова (λ).

Пример. Дан ряд значений момента трения, Н·м, фрикционного дифференциала для постоянного значения передаваемого момента:

202, 198, 193, 206, 212, 201, 196, 190, 197, 209, 203, 207, 196, 213, 203.

Расчёт статистических характеристик приводит к следующим результатам:

$\bar{x} = m_1 = 201{,}73$; $D = 43{,}39$; $\sigma = 6{,}587$;

$D_0 = 46{,}49$; $\sigma_0 = 6{,}819$; $A = 0{,}06997$;

$E = -0{,}7849$; $m_3 = 20{,}001$; $m_4 = 4171{,}263$;

$U_3 = 0{,}5401$; $U_4 = 0{,}8921$.

Проверка условия (2) показывает соответствие экспериментальных данных нормальному закону распределения.

4. Дисперсионный анализ.

В научных исследованиях часто требуется оценка достоверности влияния на исследуемый процесс фактора, который может быть учтён лишь качественно. Решение такой задачи обеспечивает применение дисперсионного анализа.

Метод дисперсионного анализа предусматривает (для случая однофакторного эксперимента) сравнение двух серий измерений какого-либо параметра одного и того же процесса. Одна из серий принимается в качестве базовой, другая получена в условиях действия исследуемого фактора.

По данным сравниваемых серий измерений рассчитывается сумма квадратов отклонений от общего среднего между сериями Q_1 и внутри серии Q_2.

$$Q_1 = n \cdot \sum_{i=1}^{m} (\bar{x}_i - \bar{x})^2 ; \quad Q_2 = \sum_{i=1}^{m} \sum_{i=1}^{n} (x_{ij} - \bar{x})^2 ;$$

где $\bar{x}_i$ - среднее арифметическое для n измерений в серии; $\bar{x}$ - среднее арифметическое для обоих серий (общее среднее значение); x_{ij} - отдельное измерение в j-той серии; m - число сравниваемых серий. Величина Q_1 характеризует рассеивание исследуемого параметра между сериями, а величина Q_2 - рассеивание этого параметра в серии.

Оценка достоверности влияния фактора на выбранный параметр процесса производится по критерию Фишера:

$$F_p = \frac{\sigma_1^2}{\sigma_2^2};$$

где

$\sigma_1^2 = \dfrac{Q_1}{m-1}$ - дисперсия между сериями, обусловленная изучаемым фактором;

$\sigma_2^2 = \dfrac{Q_2}{m \cdot (n-1)}$ - дисперсия внутри серии, обусловленная случайными причинами.

Если изучаемый фактор оказывает существенное влияние на изучаемый параметр, то:

$$\sigma_1^2 >> \sigma_2^2 .$$

Чем больше величина критерия Фишера F, тем больше степень достоверности этого влияния. Для проверки статистической значимости исследуемого влияния расчётное значение критерия F_p сравнивают с табличным F_T, выбираемым в зависимости от числа степеней свободы $f_1 = m-1$, $f_2 = m \cdot (n-1)$ при данной доверительной вероятности $P_Д = 0{,}95$.

Если $F_P > F_T$, то влияние фактора на исследуемый параметр значимо. Табличные значения критерия Фишера приведены в табл. 10.

Пример. Оценить значимость влияния замены упругого элемента в подвеске легкового автомобиля на частоту колебаний подрессоренной массы автомобиля Ока-1113 при исходных данных, приведенных в табл. 8.

Табл. 8.
Исходные данные.

№ п/п	*Частота колебаний Ω, Гц*	
	Базовый упругий элемент	*Опытный упругий элемент*
1	*1,05*	*1,06*
2	*1,08*	*1,07*
3	*0,95*	*1,04*
4	*0,98*	*1,02*
5	*1,01*	*1,08*
6	*0,97*	*1,03*

Промежуточные выкладки приведены в табл. 9.

Табл. 9.
Данные расчёта.

№ п/п	x_A	x_B	$x_A-\bar{x}_A$	$(x_A-\bar{x}_A)^2$	$x_B-\bar{x}_B$	$(x_B-\bar{x}_B)^2$
1	1,05	1,06	0,0433	0,00187	0,01	0,0001
2	1,08	1,07	0,0733	0,00537	0,02	0,0004
3	0,95	1,04	-0,0567	0,00321	-0,01	0,0001
4	0,98	1,02	-0,0267	0,000713	-0,03	0,0009
5	1,01	1,08	0,0033	0,0000109	0,03	0,0009
6	0,97	1,03	-0,0367	0,00135	-0,02	0,0004
Σ	6,04	6,3		0,0125239		0,0028

$\bar{x}_A = 1{,}0067$;

$\bar{x}_B = 1{,}05$;

$\bar{x} = 1{,}028$;

$Q_1 = 6\cdot((1{,}0067-1{,}028)^2+(1{,}05-1{,}028)^2) = 0{,}00563$;

$Q_2 = 0{,}01252 + 0{,}0028 = 0{,}01532$.

Определяем дисперсии между сериями и внутри серий.

$$\sigma_1^2 = \frac{0{,}00563}{2-1} = 0{,}00563\,; \quad \sigma_2^2 = \frac{0{,}01532}{2\cdot(6-1)} = 0{,}001532\,.$$

Расчётное значение критерия Фишера:

$$F_P = \frac{0{,}00563}{0{,}001532} = 3{,}675\,.$$

По табл. 10 для известных $P_Д = 0{,}95$, $f_1 = 2-1 = 1$ *и* $f_2 = 2\cdot(6-1) = 10$ *находим значение* $F_T = 4{,}96$. *Так как* $F_P < F_T$, *то влияние замены упругого элемента на частоту колебаний подрессоренной массы не является значимым.*

Табл. 10.
Критерий Фишера.

f_2	Значение F_T при $P_Д = 0{,}95$ и различных f_1								
	1	2	3	4	5	6	7	8	9
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,39
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,69
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49
18	4,41	3,55	3,!6	2,93	2,77	2,66	2,58	2,51	2,46
19	4,38	3,52	3,13	2,89	2,74	2,63	2,54	2,48	2,42
20	4,35	3,49	3,!0	2,87	2,71	2,60	2,51	2,45	2,39

5. Расчёт характеристик парной корреляции. Линейный парный регрессионный анализ.

Корреляция является признаком, указывающим на взаимосвязь ряда численных последовательностей. Парная корреляция характеризует взаимосвязь двух последовательностей x_i и y_i. Постоянные коэффициенты a и b в линейном уравнении регрессии типа $y = a \cdot x + b$ по методу наименьших квадратов рассчитываются по формулам:

$$b = \frac{n \cdot \sum_{i=1}^{n} x_i \cdot y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{n \cdot \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}; \quad a = y - b \cdot x = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i - \frac{b}{n} \cdot \sum_{i=1}^{n} x_i;$$

где n - число парных измерений величин x и y.

Коэффициент парной корреляции находится по формуле:

$$R = \frac{\sum_{i=1}^{n} x_i \cdot y_i - \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}} \cdot \sqrt{\sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}}}$$

Он характеризует степень отклонения связи между x_i и y_i от линейной. Если $|R|$ близок к 1, то эта связь линейна, т.е. $y_i = a \cdot x_i + b$, причём знак R определяет знак коэффициента a. Если $R > 0$, то $a > 0$, и, напротив, при $R < 0$ $a < 0$. Величина коэффициента корреляции указывает на тесноту связи между переменными. Если $|R| = 1{,}0$, то связь является чисто линейной; если $|R| = 0$, то корреляционной связи между x_i и y_i нет или она нелинейная. Обычно считают тесноту связи удовлетворительной при $|R| \geq 0{,}5$, линейность связи достоверной при $|R| \geq 0{,}95$.

При малом числе парных измерений ($n \leq 50$) оценивается достоверность связи путём расчёта дисперсии коэффициента корреляции по уравнению:

$$\sigma_r = \frac{1 - R^2}{\sqrt{n-1}}.$$

Связь считается достоверной с доверительной вероятностью $0{,}95$, если выполняется условие:

$$\frac{|R|}{\sigma_r} \geq 3.$$

При нелинейной корреляционной связи обычно используют уравнение в виде многочлена:

$$y = a + b \cdot x + c \cdot x^2.$$

Пример. Методом наименьших квадратов определить коэффициенты линейного уравнения, тесноту и достоверность связи между величиной макронеровностей дороги $h_д$ и изменением максимальной динамической реакции на упругом элементе подвески $\Delta R_{z\,max}$ при следующих данных:

Табл. 11.
Исходные данные.

$h_д$ (x), мм	16	27	36	43	53	64
$\Delta R_{z\,max}$ (y), Н	438	522	629	745	880	1035

Заносим исходные значения x, y и промежуточные расчётные величины в сводную таблицу.

Табл. 12.
Сводные данные.

№ п/п	x	y	$x \cdot y$	x^2	y^2
1	16	438	7008	256	191844
2	27	522	14094	729	272484
3	36	629	22644	1296	395641
4	43	745	32035	1849	555025
5	53	880	46640	2809	774400
6	64	1035	66240	4096	1071225
Σ	239	4249	188661	11035	3260619

Определим значения постоянных коэффициентов в линейном уравнении регрессии:

$$b = \frac{6 \cdot 188661 - 239 \cdot 4249}{6 \cdot 11035 - 239^2} = \frac{116455}{9089} = 12{,}813$$

$$a = \frac{4249}{6} - 12{,}813 \cdot \frac{239}{6} = 708{,}167 - 510{,}385 = 197{,}783 .$$

Таким образом, уравнение линейной связи имеет вид:

$$\Delta R_{z\,max} = 197{,}783 + 12{,}813 \cdot h_{\partial} .$$

Определим коэффициент корреляции:

$$R = \frac{6 \cdot 188661 - 239 \cdot 4249}{\sqrt{(6 \cdot 11035 - 239^2) \cdot (6 \cdot 3260619 - 4249^2)}} =$$

$$= \frac{116455}{\sqrt{9089 \cdot 1509713}} = \frac{116455}{117140} = 0{,}9942 .$$

Полученное значение R, близкое к 1, свидетельствует о чисто линейном характере связи. В связи с малым числом измерений проверяем достоверность связи расчётом дисперсии коэффициента корреляции:

$$\sigma_r = \frac{1 - 0{,}9942^2}{\sqrt{6-1}} = \frac{0{,}01157}{2{,}236} = 0{,}00517 ;$$

$$\frac{R}{\sigma_r} = \frac{0{,}9942}{0{,}00517} = 192{,}302 >> 3 .$$

Связь достоверна.

Линейный парный регрессионный анализ заключается в определении параметров эмпирической линейной зависимости:

$$y(x) = b_1 \cdot x + b_0 ; \qquad (3)$$

описывающей связь между некоторым числом n пар значений x_i и y_i, обеспечивая при этом наименьшую среднеквадратичную погрешность. Графически эту задачу можно представить следующим образом - в облаке точек $\{x_i ; y_i\}$ плоскости XOY требуется провести прямую так, чтобы величина всех отклонений отвечала условию:

$$U = \sum_{i=1}^{n} [\, y_i - y(x_i)]^2 = min ,$$

где $y(x_i)$ - зависимость (3). Для этого нужно приравнять нулю частные производные:

$$\frac{\partial U}{\partial b_0} = \sum_{i=1}^{n} [\, y_i - (b_0 + b_1 \cdot x_i)] ; \quad \frac{\partial U}{\partial b_1} = \sum_{i=1}^{n} [\, y_i - (b_0 + b_1 \cdot x_i) \cdot x_i \,] ;$$

что даёт для определения неизвестных коэффициентов b_0 и b_1 систему линейных уравнений:

$$b_0 \cdot n + b_1 \cdot \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i ; \qquad b_0 \cdot \sum_{i=1}^{n} x_i + b_1 \cdot \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i \cdot y_i .$$

Решение этой системы:

$$b_1 = \frac{\sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i - n \cdot \sum_{i=1}^{n} x_i \cdot y_i}{\left(\sum_{i=1}^{n} x_i\right)^2 - n \cdot \sum_{i=1}^{n} x_i^2} ; \quad b_0 = \frac{1}{n} \cdot \left(\sum_{i=1}^{n} y_i - b_1 \cdot \sum_{i=1}^{n} x_i \right).$$

При необходимости можно произвести дополнительно вычисление коэффициента парной корреляции R или среднеквадратичной погрешности:

$$\sigma^2 = \frac{1}{n} \cdot \left[\sum_{i=1}^{n} y_i^2 - b_1 \cdot \sum_{i=1}^{n} y_i - b_0 \cdot \sum_{i=1}^{n} x_i \cdot y_i \right],$$

что позволит количественно оценивать степень приближения точек $\{x_i ; y_i\}$ к прямой.

Нелинейная парная регрессия сводится к получению заданной нелинейной зависимости $y(x)$ (нелинейной по независимой переменной x, но линейной по параметрам этой зависимости!), приближающей совокупность чисел x_i и y_i с наименьшей среднеквадратичной погрешностью.

Сведение нелинейной регрессии к линейной выполняется с помощью линеаризующих преобразований в ходе ввода x_i, y_i и при выводе b_0 и b_1. Такое сведение целесообразно, если в распоряжении имеется только программа вычисления параметров линейной регрессии.

6. Нелинейный регрессионный анализ.

В общем случае для нелинейной регрессии более удобно использовать методики и программы, не требующие специальных преобразований при вводе x_i, y_i и при выводе результатов вычислений.

Гиперболическая регрессия заключается в нахождении параметров функции:

$$y(x) = b_0 + b_1 / x;$$

из решения системы уравнений

$$b_0 \cdot n + b_1 \cdot \sum_{i=1}^{n} 1/x_i = \sum_{i=1}^{n} y_i ; \; b_0 \cdot \sum_{i=1}^{n} 1/x_i + b_1 \cdot \sum_{i=1}^{n} 1/x_i^2 = \sum_{i=1}^{n} y_i / x_i .$$

Степенная регрессия обеспечивает нахождение параметров функции:

$$y = b_0 \cdot x^{b_1};$$

по формулам

$$b_1 = \frac{\sum_{i=1}^{n} \ln x_i \cdot \sum_{i=1}^{n} \ln y_i - n \cdot \sum_{i=1}^{n} \ln x_i \cdot \ln y_i}{\left(\sum_{i=1}^{n} \ln x_i\right)^2 - n \cdot \sum_{i=1}^{n} (\ln x_i)^2};$$

$$b_0 = \exp\left[\frac{1}{n} \cdot \left(\sum_{i=1}^{n} \ln y_i - b_1 \cdot \sum_{i=1}^{n} \ln x_i\right)\right].$$

Показательная регрессия обеспечивает получение параметров a и b показательной функции:

$$y(x) = a \cdot b^x;$$

из решения системы уравнений

$$n \cdot \lg a + \lg b \cdot \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \lg y_i;$$

$$\lg a \cdot \sum_{i=1}^{n} x_i + \lg b \cdot \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} (x_i \cdot \lg y_i).$$

Экспоненциальная регрессия обеспечивает получение параметров функции:

$$y = b_0 \cdot \exp(b_1 \cdot x);$$

по формулам

$$b_1 = \frac{\sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} \ln y_i - n \cdot \sum_{i=1}^{n} x_i \cdot \ln y_i}{\left(\sum_{i=1}^{n} x_i\right)^2 - n \cdot \sum_{i=1}^{n} x_i^2};$$

$$b_0 = \exp\left[\frac{1}{n} \cdot \left(\sum_{i=1}^{n} \ln y_i - b_1 \cdot \sum_{i=1}^{n} x_i\right)\right].$$

Логарифмическая регрессия даёт параметры a и b функции:

$$y = a + b \cdot lg\, x;$$

из решения системы уравнений

$$a \cdot n + b \cdot \sum_{i=1}^{n} lg\, x_i = \sum_{i=1}^{n} y_i ;$$

$$a \cdot \sum_{i=1}^{n} lg\, x_i + b \cdot \sum_{i=1}^{n} (lg\, x_i)^2 = \sum_{i=1}^{n} (y_i \cdot lg\, x_i).$$

Параболическая регрессия обеспечивает получение трёх параметров b_0, b_1 и b_2 приближения параболической функцией:

$$y = b_0 + b_1 \cdot x + b_2 \cdot x^2 ;$$

из решения системы уравнений

$$b_0 \cdot n + b_1 \cdot \sum_{i=1}^{n} x_i + b_2 \cdot \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i ;$$

$$b_0 \cdot \sum_{i=1}^{n} x_i + b_1 \cdot \sum_{i=1}^{n} x_i^2 + b_2 \cdot \sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} x_i \cdot y_i ;$$

$$b_0 \cdot \sum_{i=1}^{n} x_i^2 + b_1 \cdot \sum_{i=1}^{n} x_i^3 + b_2 \cdot \sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} x_I^2 \cdot y_i .$$

Полиномиальная регрессия (аппроксимация) обеспечивает нахождение коэффициентов полинома:

$$y(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2 + ... + a_m \cdot x^m ;$$

из решения системы уравнений

$$\begin{cases} c_0 \cdot a_0 + c_1 \cdot a_1 + c_2 \cdot a_2 + ... + c_m \cdot a_m = d_0 \\ c_1 \cdot a_0 + c_2 \cdot a_1 + c_3 \cdot a_2 + ... + c_{m+1} \cdot a_m = d_1 \\ ... \\ c_m \cdot a_0 + c_{m+1} \cdot a_1 + c_{m+2} \cdot a_2 + ... + c_{2m} \cdot a_m = d_m \end{cases} ; \qquad (4)$$

где

$$c_j = \sum_{i=1}^{n} x_i^j ,\ j = 0,1,2,...,2 \cdot m; \quad d_k = \sum_{i=1}^{n} x_i^k \cdot y_i ,\ k = 0,1,2,...,m .$$

Полученная система линейных уравнений решается методом Гаусса. Полином степени $m < n$, где n - число пар x_i, y_i, обеспечивает аппроксимацию (и интерполяцию) таблично заданной функции $y_i(x_i)$ с минимальной среднеквадратичной погрешностью:

$$E = \sqrt{\sum_{i=1}^{n} \frac{\varepsilon_i^2}{n+1}}\,. \qquad (5)$$

Если $m = n$, то имеет место обычная интерполяция, т.е. значения $y(x)$ при $x = x_i$ точно совпадают с заданными y_i. При $m < n$ такого совпадения в общем случае нет. Таким образом, аппроксимация $y_i(x_i)$ по методу наименьших квадратов имеет более универсальный характер, чем обычная интерполяция.

Полиномиальная регрессия (аппроксимация) с автоматическим выбором степени полинома выполняется по следующему алгоритму. Вначале задаётся степень $m = 1$ (линейная регрессия). Отсчёты x_i, y_i вводятся и запоминаются. После нахождения вначале всех a_0 и a_1, затем a_2, вычисляется среднеквадратичная погрешность по формуле (5) и сравнивается с заданной E_0. Если $E > E_0$, степень полинома увеличивается на *1*, и т.д. Счёт прекращается, как только достигается $E < E_0$.

При оценке достоверности приближения функции методом регрессионного анализа следует определить среднее значение относительной погрешности, %:

$$\bar{\delta} = \frac{1}{n} \cdot \sum_{i=1}^{n} \left| (y_{эi} - y_{pi}) / y_{pi} \right| \cdot 100\,; \qquad (6)$$

где $y_{эi}$ - экспериментальные значения функции, y_{pi} - значения функции, полученные с помощью регрессионной зависимости; n - число опытов (пар значений). Расхождение данных $\geq 15\%$ свидетельствует о неточности применяемого метода. В таком случае необходимо сменить метод регрессии или увеличить степень полинома (для полиномиальной регрессии). Расхождение в пределах

$5...15\%$ указывает на удовлетворительную точность метода, в пределах $\leq 5\%$ - на высокую точность метода.

Пример. Определить регрессионные зависимости для пар значений (x, y), приведенных в табл. 7, и соответствующие погрешности.

Применяется полиномиальная регрессия первого порядка $(m = 1)$. Решение системы уравнений (4) даёт следующие формулы аппроксимации:

Для y_i:

$$y = 970{,}45 - 0{,}727 \cdot x.$$

Для $\bar{y}_i$ *(сглаживание по трём точкам):*

$$y = 970{,}63 - 0{,}730 \cdot x.$$

Для $\bar{y}_i$ *(сглаживание по пяти точкам):*

$$y = 975{,}44 - 0{,}766 \cdot x.$$

Значения y, *полученные по регрессионным зависимостям, и соответствующие относительные погрешности приведены в табл. 13.*

Табл. 13.
Результаты аппроксимации.

i	*Значения регрессионных функций*			*Точное значение* y
	y_i	$\bar{y}_i$ *(сглажена по трём точкам)*	$\bar{y}_i$ *(сглажена по пяти точкам)*	
0	970,45	970,63	975,44	977,0
1	955,91	956,03	960,11	959,78
2	941,36	941,43	944,78	942,56
3	926,82	926,83	929,45	925,34
4	912,27	912,23	914,12	908,12
i	*Значения регрессионных функций*			*Точное*

	y_i	$\bar{y}_i$ *(сглажена по трём точкам)*	$\bar{y}_i$ *(сглажена по пяти точкам)*	*значение* y
5	*897,73*	*897,63*	*898,79*	*890,9*
6	*883,18*	*883,03*	*883,47*	*873,68*
7	*868,63*	*868,43*	*868,14*	*856,46*
8	*854,09*	*853,83*	*852,81*	*839,24*
9	*839,55*	*839,23*	*837,48*	*822,02*
$\bar{\delta}$ *,%*	*0,889*	*0,873*	*0,829*	-

Можно сделать вывод о высокой точности применяемого метода и достаточной эффективности проведенного ранее сглаживания данных эксперимента.

7. Аппроксимация и интерполяция с помощью сплайн-функций.

Полиномиальная интерполяция и аппроксимация не обеспечивает непрерывность производных функции $y(x)$ и может давать значительные погрешности в промежутках между узлами (расчётными точками). Кроме того, она плохо приспособлена для экстраполяции и, как правило, не обеспечивает правильное асимптотическое поведение $y(x)$ при изменении аргумента x за пределами интервала интерполяции. Нередко с увеличением числа узлов погрешность такой интерполяции не только не уменьшается, но и начинает расти.

От этих недостатков свободна аппроксимация и интерполяция с помощью сплайн-функций. Сплайн-функцию (англ. *Splain* - *гибкая линейка*) можно наглядно трактовать как линию, которую образует гибкая линейка, будучи закреплённой в ряде точек - узлах интерполяции. Математически сплайн - специальный многочлен, принимающий в узлах функции значения $y(x) = y_i = y(x_i)$ и обеспечивающий непрерывность в них производных. Обычно достаточно обеспечить непрерывность первой и второй производных, для чего достаточно использовать сплайн-многочлены третьего порядка (кубические сплайны).

Для каждого отрезка $[x_i, x_{i+1}]$ изменения x кубическая сплайн-функция записывается в виде:

$$f_i(x) = \frac{1}{6 \cdot h_i} \cdot [m_i \cdot (x_{i+1} - x)^3 + m_{i+1} \cdot (x - x_i)^3] + \frac{1}{h_i} \cdot$$

$$\cdot [(y_i - \frac{m_i \cdot h_i^2}{6}) \cdot (x_{i+1} - x) + (y_{i+1} - \frac{m_{i+1} \cdot h_i^2}{6}) \cdot (x - x_i)];$$

$$h_i = x_{i+1} - x_i; \; f_i(x) = y(x); \; m_i = f''(x_i); \; i = 1,2,...,n;$$

где n - число узлов. При известных x_i, y_i и m_i эта формула задаёт сплайн-аппроксимацию.

Если потребовать выполнения условия $f_i(x) = y_i$, то приведённое выше выражение для кубических сплайн-полиномов приводит к системе линейных уравнений, из которых находится m_i:

$$h_i \cdot m_i + 2 \cdot (h_i + h_{i+1}) \cdot m_{i+1} + h_{i+1} \cdot m_{i+2} =$$

$$= 6 \cdot \left(\frac{y_{i+2} - y_{i+1}}{h_{i+1}} - \frac{y_{i+1} - y_i}{h_i} \right).$$

Однако эта система не полностью определяет m_i. Чтобы полностью определить все m_i, нужно задать дополнительные граничные условия. Если они заданы в виде $m_1 = 0$ и $m_n = 0$, получаем нормальные сплайн-функции; при $m_n = m_1$ и $m_{n+1} = m_2$ имеем периодические сплайн-функции и т.д.

По данному алгоритму необходимо вычислить коэффициенты $m_i = f''(x_i)$ нормального кубического сплайна, что при $x \in [x_1, x_n]$ обеспечит интерполяцию с помощью описанной выше сплайн-функции. Если $x < x_1$, то выполняется линейная экстраполяция по формуле:

$$f(x) = y_1 - ((x_2 - x_1) \cdot m_2 / 6 + (y_2 - y_1)/(x_2 - x_1)) \cdot (x - x_1);$$

а при $x > x_n$ - экстраполяция по формуле:

$$f(x) = y_n + ((x_n - x_{n-1}) \cdot m_{n-1}/6 + (y_n - y_{n-1})/(x_n - x_{n-1})) \cdot$$
$$\cdot (x - x_n).$$

При экстраполяции предполагается, что наклон линейного участка $f(x)$ равен первой производной сплайн-функции в точке (x_1, y_1) при экстраполяции для $x < x_1$, и в точке (x_n, y_n) при $x > x_n$.

Сложные колебательные процессы аппроксимируются периодическими сплайн-функциями. Устранение осцилляций при обработке экспериментальных данных, обусловленных наличием случайных ошибок и промахов испытаний, может быть получено с помощью сглаживающих сплайнов. Оценку достоверности влияния фактора следует производить по критерию Фишера. В случае взаимного влияния двух параметров и сложной выходной характеристики возможно применение бикубических сплайнов для решения задачи аппроксимации функции двух переменных. Погрешность сплайн-аппроксимации может быть также определена по формуле (6).

8. Оптимизация в инженерных задачах.

Для использования методов теории оптимизации при решении конкретной инженерной задачи необходимо установить границы подлежащей оптимизации инженерной системы, определить количественный критерий, на основе которого можно произвести анализ вариантов, и, наконец, построить модель, отражающую взаимосвязи между переменными. Эта последовательность действий составляет содержание процесса постановки задачи инженерной оптимизации.

Показатель, которым оценивается принимаемое решение, является критерий оптимальности данной задачи, а функция, выражающая значение критерия через управляемые параметры, есть целевая функция Y.

Пример. Рассмотрим задачу оптимизации конструктивных параметров на примере вала ведущей цилиндрической шестерни главной передачи автомобиля КамАЗ.

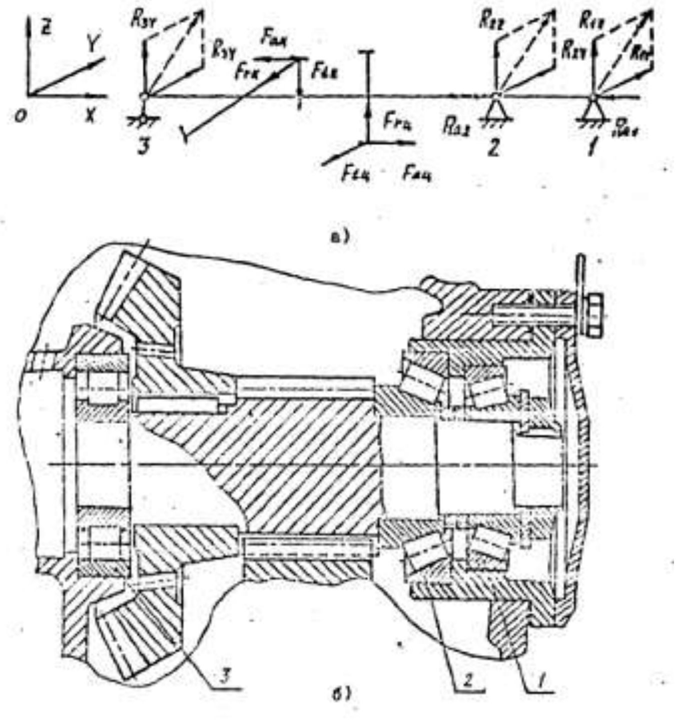

Рис. 1. Схема нагружения системой пространственных сил (а) и конструкция (б) вала ведущей цилиндрической шестерни главной передачи автомобиля КамАЗ: 1 - подшипник 27310, 2 - подшипник 27911, 3 – подшипник 102409.

В предложенной для рассмотрения задаче управляемыми (искомыми) параметрами подшипникового узла главной передачи автомобиля, определяющими его конструктивный вариант, на основании приведенных выше результатов исследований можно принять:

- углы контакта x_1 *и* x_2 *конических роликоподшипников подшипникового узла;*

- расстояние между элементами подшипникового узла x_3*;*

- усилие преднатяга x_4 *подшипникового узла при сборке;*

- динамическая вязкость масла x_5.

Учет большего числа управляемых параметров приведет к сильному усложнению и удорожанию процесса проектирования, поэтому можно ограничиться вышеприведенными.

Считая неуправляемые переменные постоянными для данной задачи, примем в качестве целевой функции функцию Y(X), которая каждому фиксированному значению набора искомых параметров ставит в соответствие некоторое определенное значение набора технических показателей.

Эта задача является задачей оптимизации, ее решение классическими методами затруднено, и поэтому требуется применение методов математического программирования и современных вычислительных средств.

Для достижения поставленной задачи большое значение имеет правильный выбор критерия оптимальности. Для этого необходимо определить частные критерии подшипникового узла главной передачи, причем при этом необходимо также учитывать влияние параметров узла на работоспособность зубчатых передач. На основании исследования данных дефектов главных передач автомобилей КамАЗ можно сформулировать следующие частные критерии оптимальности конструкции:

- нормальные силы F_{r1} *и* F_{r2} *на роликах подшипников 1 и 2;*

- зазор δ *в в подшипниковом узле в конце эксплуатации (или в определенный ее момент);*

- осевые f_a *и радиальные* f_r *перемещения зубчатых колес главной передачи.*

Теперь необходимо определить составляющие целевой функции. Как отмечалось выше: оптимальный выбор конструктивных параметров главной передачи автомобиля представляется возможным при наличии количественных оценок работоспособности этого узла.

Для проведения рационального выбора конструктивных параметров необходимо разработать целевую функцию Y. Целевая функция образуется по аддитивному принципу и представляет собой сумму произведений нормированных частных критериев на их весовые коэффициенты:

$$Y = \sum_{i=1}^{n} \alpha_i \cdot \left(X_i / X_{oi} \right),$$

при $\sum_{i=1}^{n} \alpha_i = 1$

где X_i - *численное значение* i*-го критерии;*

X_{oi} - *норма* i*-го критерия;*

α - *весовой коэффициент;*

n - *число критериев.*

В качестве нормы используется значение i*-го критерия для базовой главной передачи или его нормативная величина.*

Подиновский В.В. в работе [14] указывает, что количественная форма представления информации о важности критериев является сложной и потому зачастую ненадежной. Более доступной и достоверной, чем количественная, является качественная информация, состоящая из сведений типа «критерии равноценны», «один критерий важнее другого» и т.п. В таких случаях для учета различной значимости критериев проводят ранжирование их по важности.

В результате ранжирования критериев по важности к первому рангу можно отнести максимальный в зазор в конце эксплуатации в конических роликоподшипниках: ко второму - осевые а и радиальные а перемещения зубчатых колес конической пары: к третьему - нормальные силы А и А на роликах подшипников 1 и 2. Ранги переводятся в весовые коэффициенты по формулам без участия экспертов:

$$\alpha_i^{`} = 1 - (i - 1) / \mathrm{n},$$

При этом считается, что критерию присвоен номер назначенного ему ранга. Нормирование $\alpha_i^`$ для перехода к α_i, удовлетворяющих условию $\sum_{i=1}^{n}\alpha_i = 1$, не представляет труда.

$$\alpha_i = \frac{\alpha_i^`}{\sum \alpha_i^`}.$$

В качестве задания необходимо определить весовые коэффициенты и записать целевую функцию в общем виде.

Рекомендуемая литература

1. Автомобили: Конструкция, конструирование и расчёт. Трансмиссия. Учебное пособие для спец. “Автомобили и тракторы”. // А.И. Гришкевич, В.А. Вавуло, А.В. Карпов и др. Под ред. А.И. Гришкевича. – Мн.: Выш. шк., 1985. – 240 с., ил.
2. Ашмарин И.А., Васильнев Н.Н., Амбросов В.А. Быстрые методы статистической обработки и планирования экспериментов. - Л.: Изд. Ленинградского университета, 1975. - 47 с.
3. Баптизманский В.И. и др. Основы научных исследований. - М.: Наука, 1985.
4. Барыкин А.Ю. Основы теории современных дифференциалов. - Наб. Челны: КамПИ, 2001. – 277 с: ил.
5. Барыкин А.Ю. Технико-экономический синтез трансмиссий. - Наб. Челны: КамПИ, 2003. - 120 с., ил., табл.
6. Бендат Дж., Пирсол А. Прикладной анализ случайных данных. – М.: Мир, 1989.
7. Бахмутов С.В., Безверхий С.Ф. Статистическая обработка результатов и планирование эксперимента при испытаниях автомобиля. – М.: МГААТМ, 1994. – 86 с.
8. Гольд Б.В. и др. Прочность и долговечность автомобиля. – М.: Машиностроение, 1974.
9. Гришкевич А.И. Автомобили: теория. Учебник для вузов. - Мн.: Высш. шк., 1986. - 208 с.: ил.

10. Дьяконов В.П. Справочник по алгоритмам и программам на языке Бейсик для персональных ЭВМ. - М.: Наука, 1987. - 240 с.
11. Корн Г., Корн Т. Справочник по математике для научных работников и инженеров. - М.: Наука, 1984.
12. Мальцев Ю.А. Основы научных исследований: Учебное пособие. – М.: Балашиха, ВТУ, 2003. – 277 с.
13. Организация эксперимента. // Методические указания по практическим занятиям. Сост. Колесников М.С., Марданова М.М. - Наб. Челны: КамПИ, 2003. - 40 с.
14. Подиновский В.В., Гаврилов В.М. Оптимизация по последовательно применяемым критериям. - М.: Сов. радио, 1975. – 192 с.
15. Шелопаев С.И. Математические методы моделирования. Учебное пособие. – М.: ЮНИТИ, 2001. – 367 с.

ОГЛАВЛЕНИЕ

Учебное издание

Алексей Юрьевич Барыкин, Руслан Рамилевич Басыров,
Мансур Миргарифанович Мухаметдинов

АНАЛИЗ ДАННЫХ ЭКСПЕРИМЕНТА И ОПТИМИЗАЦИЯ В ТЕХНИЧЕСКИХ ЗАДАЧАХ

Компьютерный набор и верстка: А.Ю. Барыкин, Р.Р. Басыров, М.М. Мухаметдинов

Печатается в авторской редакции

Отпечатано с готового оригинал-макета
в Издательско-полиграфическом центре Набережночелнинского
института Казанского (Приволжского) федерального
университета

Подписано в печать 15.02.18 г. Формат 60x84/16. Печать ризографическая
Бумага офсетная №1 Гарнитура «Times New Roman»
Усл. п. л. 2,21 Уч.-изд. л. 2,21 Тираж 50 экз. Заказ № 465

423810, г. Набережные Челны, Новый город, проспект Мира, 68/19
Тел./факс (8552) 39-65-99 e-mail: ic-nchi-kpfu@mail.ru