

Лекция 10

РЕГРЕССИОННЫЕ МОДЕЛИ РЕЗУЛЬТАТОВ ИЗМЕРЕНИЙ НА ВРЕМЯ

Регрессионные модели результатов измерений на время рассмотрим на примере решения следующей задачи.

С целью диагностики подшипника двигателя с периодичностью 2 раза в месяц в течение 2 лет с помощью акселерометра измеряли уровень вибрации на частоте вращения вала (3000 об/мин) в заданной полосе частот. Результаты измерений отражены в табл. Л10.1.

Таблица Л10.1. Результаты диагностики подшипника двигателя

№ изм.	L, дБ								
0	60,5	11	63,5	22	65	33	59	43	62
1	58	12	60,5	23	60	34	62	44	60
2	56	13	61,5	24	60	35	65	45	61
3	62	14	59	25	64,5	36	64	46	62
4	68	15	63,5	26	61	37	64	47	72
5	60,5	16	60	27	60	38	64	48	61
6	61,5	17	60	28	62	39	67	49	62
7	61,5	18	65,5	29	60	40	62,5	50	63
8	60,5	19	61,5	30	64	41	64		
9	60	20	60,5	31	66	42	67		
10	62	21	60,5	32	62	43	62		

На основе приведенных данных необходимо:

1. Найти аналитическую зависимость, наиболее точно описывающую детерминированную составляющую приведенного временного ряда (построить функцию регрессии уровня вибрации на время).
2. Проверить адекватность построенной модели реальному физическому процессу.
3. Определить наличие или отсутствие значимого тренда уровня вибрации.
4. Оценить остаточный ресурс подшипника, исходя из допустимого повышения уровня вибрации на 6 дБ.

Обратим внимание на сравнительно большой объем выборки, что позволяет использовать критерии, основанные на гипотезе о нормальности распределения, которая, естественно подлежит предварительной проверке. В данном случае такая проверка затруднена, поскольку в процессе набора данных (в течение 2 лет) могли произойти изменения параметров распределения, которые и подлежат выявлению.

Поэтому первым этапом является установление наличия тренда, для чего временной ряд, отраженный в табл. Л10.1 ($N = 51$), разбивают на две части с числом точек N_1 и N_2 . При отсутствии значимого тренда полученные выборки будут принадлежать одной генеральной совокупности, что принимается в качестве нулевой гипотезы. Разобьем данные на две выборки с номерами 1...25 ($N_1 = 25$) и 26...50 ($N_2 = 25$). Один из результатов отбрасываем, как не образующий пары. Отброшен результат с нулевым номером, поскольку в реальных условиях на первые результаты измерений в большей степени, чем на последующие, влияют «приработка» оборудования, персонал и другие факторы, хотя подобный выбор может обсуждаться дополнительно.

Если принять за положительную разность между значениями в первой и второй выборках, то получим $n_1 = 24$; $k = 6$; $m = 18$. Из статистических таблиц следует, что с вероятностью ошибки 5% выборки следует считать различными, поскольку допустимая область — [7, 17]. Заметим, что для более надежного вывода (с меньшей вероятностью ошибки) нет оснований, и заключение о различии выборок, например с вероятностью ошибки 1%, не является обоснованным. Таким образом, с вероятностью ошибки не более 5% имеет место тренд уровня вибрации.

Рассмотрим в качестве первого приближения линейную модель тренда

$$L(t) = b_0 + b_1 t \quad (\text{Л10.1})$$

или

$$L(t) = a + b(t - \bar{t}), \quad (\text{Л10.2})$$

в которой время отсчитывается в условных единицах, равных номеру измерения, поскольку измерения проводились через равные промежутки времени. С использованием метода наименьших квадратов, коэффициенты регрессии находятся по формулам:

$$b_0 = \bar{L} - b_1 \bar{t};$$

$$b_1 = \frac{\sum_{i=0}^N (t_i - \bar{t})(L_i - \bar{L})}{\sum_{i=0}^N (t_i - \bar{t})^2};$$

$$b = b_1; \quad a = b_0 + b \bar{t}.$$

Расчет дает:

$$L(t) = 60,5 + 0,068 t,$$

или в виде (Л10.2)

$$L(t) = 62,2 + 0,068(t - 25)$$

Оценим дисперсию коэффициентов регрессии. Исходя из уже упомянутого предположения о нормальности распределения погрешностей, следует считать, что выборочные коэффициенты регрессии также распределены нормально, и их дисперсии могут быть вычислены по формулам:

$$\sigma^2(a) = \frac{\sigma^2(L)}{N}, \quad (\text{Л10.3})$$

$$\sigma^2(b) = \frac{\sigma^2(L)}{\sum_{i=1}^N (t_i - \bar{t})^2}. \quad (\text{Л10.4})$$

Нам необходимо получить значение дисперсии воспроизводимости $\sigma^2(L)$. Для этого разобьем выборку на ряд более мелких, в пределах которых будем полагать отсутствие значимого тренда (что тоже при необходимости может быть проверено). Возьмем, например, выборку с номерами от 21 по 30. Для нее $\bar{t} = 61,7$; $s^2(L) = 4,76$. Точно так же получим значения и для других выборок:

Таблица Л10.2. Оценка дисперсии воспроизводимости

№ изм.	1...10	11...20	21...30	31...40	41...50
$s^2(L)$	17,4	4,1	4,8	2,3	12,3
L	60,0	61,5	61,7	63,5	63,4

Теперь необходимо проверить нулевую гипотезу о том, что все m генеральных совокупностей, из которых взяты выборки, имеют равные дисперсии. Для этого проще всего применить критерий Хартли, который основан на вычислении статистики

$$F_{\max} = s_{\max}^2 / s_{\min}^2$$

и сравним ее с критическим значением (см. статистические таблицы) для выбранного уровня значимости, числа выборок m и $k = N - 1$ (N — объем выборки, который предполагается одинаковым для всех выборок). При

$$F_{\max} \leq F_{\max, \alpha}(m, k)$$

нулевую гипотезу о равенстве дисперсий выборок не отвергают, в противном случае принимают альтернативную гипотезу. В нашем случае

$$F_{\max} = 17,4 / 2,3 = 7,6 > F_{\max; 0,05}(5; 9) = 7,1,$$

т. е. дисперсии не равны. Из приведенной таблицы следует, что при отбрасывании первой выборки (по изложенным ранее соображениям)

$$F_{\max} = 12,3/2,3 = 5,3 > F_{\max; 0,05}(4; 9) = 6,3,$$

и дисперсии можно считать одинаковыми и равными

$$s_{\text{воспр}}^2(L) = s^2(L) = (4,1 + 4,8 + 2,8 + 12,3)/4 = 5,9.$$

В соответствии с формулами (Л10.3) и (Л10.4) имеем:

$$s_a^2 = 0,12; \quad s_b^2 = 0,000534; \quad s_a = 0,34; \quad s_b = 0,023.$$

Доверительные интервалы для значений генеральных коэффициентов регрессии $a_{\text{ген}}$ и $b_{\text{ген}}$ определяют на основании этих оценок в соответствии с формулами

$$a - t_{\alpha, k} s_a \leq a_{\text{ген}} \leq a + t_{\alpha, k} s_a;$$

$$b - t_{\alpha, k} s_b \leq b_{\text{ген}} \leq b + t_{\alpha, k} s_b,$$

где $t_{\alpha, k}$ — значение квантили t -распределения Стьюдента уровня $P = 1 - \alpha/2$ для числа степеней свободы, равного разности между числом измерений и количеством связей, наложенных вычислениями коэффициентов a и b , т. е. $k = N - 2 = 48$.

Выбрав $\alpha = 0,05$, т.е. вероятность попадания коэффициентов в доверительный интервал, равной 0,95, и учитывая, что $t_{\alpha, k} = z_{1-\alpha/2}$ при $k > 30$ ($z_{1-\alpha/2}$ — квантиль нормального нормированного распределения), в соответствии со статистическими таблицами имеем: $z_{0,975} = 1,96$ и

$$62,2 - 1,96 \times 0,34 \leq a_{\text{ген}} \leq 62,2 + 1,96 \times 0,34 \text{ или } 61,5 \leq a_{\text{ген}} \leq 62,9;$$

$$0,068 - 1,96 \times 0,023 \leq b_{\text{ген}} \leq 0,068 + 1,96 \times 0,023 \text{ или } 0,0230 \leq b_{\text{ген}} \leq 0,113.$$

Дисперсия оценки в соответствии с выбранной моделью определяется рассеянием результатов измерений относительно линии регрессии и равна

$$s_{\text{ост}}^2(L) = s_L^2 = \frac{1}{N-2} \sum_{i=1}^N (L_i - L_{i, \text{расч}})^2. \quad (\text{Л10.5})$$

Число степеней свободы принимается равным $N - 2$, поскольку две степени свободы потеряны при определении двух коэффициентов регрессии. Кроме того, погрешность определения коэффициентов b приводит к погрешности, связанной с зоной неопределенности угла наклона прямой и расстоянием экспериментальной точки от центра распределения:

$$s_{\text{поворота}}^2(L_i) = \left\{ (t_i - \bar{t})^2 / \sum_{k=1}^N (t_k - \bar{t})^2 \right\} s^2(L).$$

С увеличением разности $t_k - \bar{t}$ дисперсия возрастает и доверительные границы расходятся в виде ветвей гиперболы (рис. Л10.1).

При вычислении доверительных интервалов для конкретных результатов измерений следует учитывать рассеяние единичных измерений около линии регрессии, поэтому суммарная дисперсия составит

$$s_{\Sigma}^2(L_i) = s^2(L) \left(1/N + (t_i - \bar{t})^2 / \sum_{k=1}^N (t_k - \bar{t})^2 \right). \quad (\text{Л10.6})$$

Доверительный интервал для значения L_i , прогнозируемого в момент времени t_i :

$$L_i = L_{i, \text{расч}} \pm t_{\alpha, i-2} s_{\Sigma}^2(L_i). \quad (\text{Л10.7})$$

Проверим адекватность полученной модели, сопоставив остаточную дисперсию (Л10.5) и дисперсию воспроизводимости. Расчет дает значение $s_{\text{ост}}^2(L) = s_L^2 = 329,9/48 = 6,87$. Ранее получено $s_{\text{воспр}}^2(L) = s^2(L) = 5,9$. Получаем $F = s_{\text{ост}}^2(L) / s_{\text{воспр}}^2(L) = 1,16 < F_{0,01}(46; 48) \approx 1,8$, т.е. различие незначимо. Число степеней свободы для выборки с большей дисперсией равно $N - 4$, т.к. дисперсия воспроизводимости вычислялась по разбиению из четырех выборок и при вычислениях на экспериментальные данные наложены четыре связи.

В общем случае следует проверить, не улучшится ли аппроксимация при увеличении степени полинома. Улучшение согласия проверяется по изменению остаточной дисперсии: если последняя уменьшается значимо, степень полинома увеличивают. При этом необходимо иметь в виду, что в оптимальном аппроксимирующем полиноме некоторая степень может отсутствовать и проверку лучше проводить, увеличивая степень на несколько единиц.

Увеличение степени сверх оптимального значения приводит к ухудшению сглаживания усредненной кривой: полином начинает «отслеживать» экспериментальные погрешности. Зону рассеяния строим согласно уравнению (Л10.6):

$$s_{\Sigma}^2(L_i) = 5,9 \left(1/50 + (t_i - 25)^2 / 11050 \right) = 0,118 + 6,22 \cdot 10^{-4} (t_i - 25)^2.$$

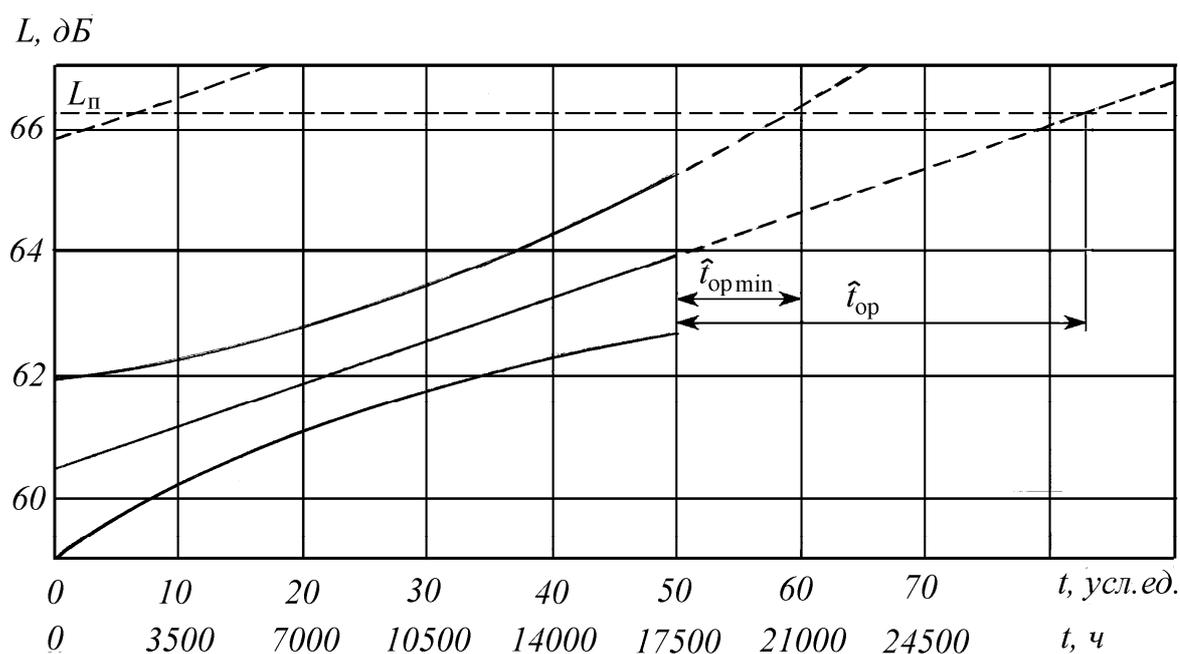


Рис. Л10.1. Пояснение расчета характеристик тренда и оценки остаточного ресурса

Построенные по этим данным зависимости отражены на рис. Л10.1. Исходя из допустимого повышения уровня на 6 дБ, получим предельно допустимое значение уровня $L_{дон} = 66,5$ дБ. Отсюда получим и оценки статочного ресурса; наиболее вероятную, соответствующую пересечению уровня $L_{дон}$ с линией тренда: $t_{оп} = 36$ усл. ед. или 12 600 ч, и минимальную, соответствующую надежности 95%, $t_{оп min} = 10$ усл. ед. или 3500 ч.

Если оценка состояния проводится по единичному измерению, то оценку (Л10.6) следует увеличить на величину погрешности единичного измерения, что приведет к выражению:

$$s_{\Sigma}^2(L_i) = s^2(L) \left(1 + 1/N + (t_i - \bar{t})^2 / \sum_{k=1}^N (t_k - \bar{t})^2 \right).$$

Соответствующая граница допуска уровня отражена пунктиром в левой верхней части рис. Л10.1. Отсюда видно, сколь статистически ненадежна оценка состояния по результатам единичного измерения.

Рассмотренный подход и иллюстрирующий его пример имели целью показать, каким образом диагностическая информация может быть использована для прогнозирования развития во времени событий на основе информации, накопленной на момент наблюдения.

Такой подход в настоящее время может считаться упрощенным, учитывая наличие большого количества вычислительных программ для обработки диагностической информации.

Рассмотрим возможные пути развития намеченного в предыдущем разделе подхода.

1. Замена простейшей линейной прогнозирующей модели более сложной моделью, что в общем случае повышает надежность прогноза. Проще всего аппроксимировать наблюдаемые значения степенным рядом, так как по степеням аргумента разлагается большинство функций. Подбирая коэффициенты к экспериментальным данным, можно аппроксимировать любую функцию. Но повышение степени полинома ведет к значительным изменениям коэффициентов при более низких степенях аргумента. Например, если описать процесс двумя моделями — $y = a + bt$ и $y = a_1 + b_1t + b_2t^2$, то в общем случае $a \neq a_1$ и $b \neq b_1$, то есть при повышении степени полинома все коэффициенты нужно пересчитывать заново. Степень улучшения качества модели по мере повышения степени полинома проверяется с помощью известных статистических критериев, например критерия Фишера.

2. В ряде случаев целесообразно применить преобразование координат с целью сведения анализа к уже рассмотренному линейному случаю. Некоторые из возможных видов преобразований указаны в табл. Л10.3.

3.

Таблица Л10.3. **Линеаризующие преобразования для проведения линейного регрессионного анализа**
($y = b_0 + b_1x \rightarrow y' = b'_0 + b'_1x'$)

Функция	Линеаризующие преобразования			
	переменных		коэффициентов регрессии	
	y'	x'	b'_0	b'_1
$y = b_0 + b_1/x$	y	$1/x$	b'_0	b_1
$y = 1/(b_0 + b_1x)$	$1/y$	x	b_0	b_1
$y = x/(b_0 + b_1x)$	x/y	x	b_0	b_1
$y = b_0 b_1^x$	$\lg y$	x	$\lg b_1$	$\lg b_1$
$y = b_0 \exp\{b_1x\}$	$\ln y$	x	$\ln b_0$	b_1
$y = 1/(b_0 + b_1 e^{-x})$	$1/y$	$\exp\{-x\}$	b_0	b_1
$y = b_0 x^{b_1}$	$\lg y$	$\lg x$	$\lg b_0$	b_1
$y = b_0 + b_1 \lg x$	y	$\lg x$	b_0	b_1
$y = b_0x/(b_1 + x)$	$1/y$	$1/x$	b_1/b_0	$1/b_0$
$y = b_0 \exp\{b_1/x\}$	$\ln y$	$1/x$	$\ln b_0$	b_1
$y = b_0 + b_1x^n$	y	x^n	b_0	b_1

4. Подгонка коэффициентов в физически обоснованной математической модели. Например, если на основе физических соображений изменение параметра должно развиваться по экспоненте, то подгонку коэффициентов следует осуществлять именно под такую модель.

В связи с этим заметим, что развитие любых физических процессов во времени может быть описано совокупностью экспонент:

- собственно экспонент вида $\exp\{at\}$, где $a \neq 0$ — действительная величина; случай $a > 0$ описывает лавинообразные процессы, $a < 0$ соответствует затухающим со временем (релаксационным) процессам;
- экспонент с мнимым показателем, являющихся периодическими функциями;
- экспонент с комплексным показателем, описывающих колебательные процессы, нарастающие или затухающие со временем;
- экспонент с нулевым показателем степени, то есть констант (более точно — скачкообразно меняющихся единичных функций).

5. Прогнозирование на основе измерения нескольких независимых параметров, описывающих состояние диагностируемой системы, что существенно улучшает надежность диагностики и прогнозирования разрушения. При этом необходимо рассмотрение движения точки, описывающей состояние системы в многомерном пространстве параметров, и вероятность выхода ее на границы их допустимых значений. Возникают дополнительные проблемы, связанные с анализом степени коррелированности (и связанной с ними обратной зависимостью степени информативности) параметров. Многомерный анализ может быть проведен только с применением достаточно мощных вычислительных средств.

Соответствующий анализ является достаточно сложной проблемой, требующей отдельного подробного рассмотрения.