ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 13.

РЕГРЕССИОННЫЙ АНАЛИЗ. ПРОВЕРКА АДЕКВАТНОСТИ РЕГРЕССИОННЫХ МОДЕЛЕЙ. ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ ДЛЯ ПАРАМЕТРОВ.

Задача 1.26. В задаче 1.25, разобранной на практическом занятии 11, рассчитаны линейная и квадратичная модели регрессии. По отдельной независимой серии измерений получена несмещенная оценка дисперсии $S^2 = 0{,}32$ с числом степеней свободы k равным 20. Проверить адекватность линейной и квадратичной моделей регрессии с уровнем значимости $\alpha = 0{,}05$.

Решение.

Все необходимые формулы представлены на практическом занятии 11. Для нахождения дисперсии адекватности (1.66) необходимо вычислить сумму квадратов отклонений результатов эксперимента от функции регрессии $\sum \Delta Y^2$ для каждой модели регрессии. Для линейной модели эта сумма равна 2,3 (см. последний столбец таблицы 1.11). Число точек, в которых проводился эксперимент, $n = 5$; число оцениваемых параметров $m = 2$, тогда $k_{ад} = 5 - 2 = 3$. Дисперсия адекватности (1.74) равна $S^2_{ад\ лин} = 2{,}3/3 = 0{,}767$. Для проверки гипотезы об адекватности линейной модели вычисляем критерий Фишера (1.67): $F = S^2_{ад} / S^2_{экс} = 0{,}767/0{,}32 = 2{,}40$. Квантиль распределения Фишера $F_{0,95}(3; 20) = 3{,}10$. Так как $2{,}4 < 3{,}1$, гипотеза об адекватности линейной модели принимается с уровнем значимости $\alpha = 0{,}05$.

Адекватность квадратичной модели проверяется аналогично. Сумма квадратов отклонений $\sum \Delta Y^2$ (см. последний столбец таблицы 1.12) равна 1,4. $n = 5$; $m = 3$; $k_{ад} = 5 - 3 = 2$; $S^2_{ад\ кв} = 1{,}4/2 = 0{,}7$, $F = S^2_{ад} / S^2_{экс} = 0{,}7/0{,}32 = 2{,}19$. $F_{0,95}(2; 20) = 3{,}49$. Так как $2{,}19 < 3{,}49$, гипотеза об адекватности квадратичной модели принимается с уровнем значимости $\alpha = 0{,}05$.

Квадратичную модель в данной задаче строить нецелесообразно. Так как линейная модель оказалась адекватной, ее уточнять не было необходимости.

Задача 1.27. В задаче 1.25, разобранной на практическом занятии 11, получено уравнение линейной и квадратичной модели регрессии. По отдельной независимой серии измерений получена несмещенная оценка дисперсии $S^2 = 0{,}32$ с числом степеней свободы k равным 20. Считая результаты экспериментов независимыми, равноточными и подчиняющимися нормальному закону распределения, построить доверительные интервалы для истинных значений коэффициентов обеих моделей с надежностью $\mathcal{P}$ = 0,95.

Решение.

Все необходимые формулы представлены на практическом занятии 11. В задаче 1.25 получена линейная модель регрессии от кодированной переменной

$$y = \widehat{B}_1 + \widehat{B}_2 X = 7,2 + 0,95X.$$

Границы доверительных интервалов (1.75) для коэффициентов B_1 и B_2 при этом будут:

$$B_1 = 7,2 \pm \varepsilon_1; \qquad \varepsilon_1 = t_{1-\frac{\alpha}{2}}(k)\frac{S}{\sqrt{n}} = t_{0,975}(20)\frac{\sqrt{0,32}}{\sqrt{5}} = 2,086\frac{\sqrt{0,32}}{\sqrt{5}} = 0,52;$$

$$B_2 = 0,95 \pm \varepsilon_2; \qquad \varepsilon_2 = t_{1-\frac{\alpha}{2}}(k)\frac{S}{\sqrt{\sum_{i=1}^{n} X_i^2}} = 2,086\frac{\sqrt{0,32}}{\sqrt{10}} = 0,38.$$

Уравнение линейной регрессии от реального переменного x

$$y = \beta_1 + \beta_2 x = 4,35 + 4,75x.$$

Границы доверительных интервалов для коэффициентов β_1 и β_2 находим по формулам (1.77):

$$\beta_1 = 4,35 \pm \tilde{\varepsilon}_1; \qquad \hat{\varepsilon}_1 = t_{1-\frac{\alpha}{2}}(k)\, S\sqrt{\frac{1}{n} + \frac{\overline{x}^2}{h^2\sum_{i=1}^{n} X_i^2}} = 2,086\sqrt{0,32}\sqrt{\frac{1}{5} + \frac{0,36}{0,04\cdot 10}} = 1,24;$$

$$\beta_2 = 4,75 \pm \hat{\varepsilon}_2; \qquad \hat{\varepsilon}_2 = t_{1-\frac{\alpha}{2}}(k)\frac{S}{h\sqrt{\sum_{i=1}^{n} X_i^2}} = 2,086\frac{\sqrt{0,32}}{0,2\sqrt{10}} = 1,9.$$

Квадратичная модель регрессии с кодированным переменным

$$y = B_1 + B_2 X + B_3 X^2 = 7,7 + 0,95X - 0,25X^2.$$

Границы доверительных интервалов (1.84) для коэффициентов B_1, B_2 и B_3:

$$B_1 = 7,7 \pm \varepsilon_1; \qquad \varepsilon_1 = t_{1-\frac{\alpha}{2}}(k)\, S\sqrt{\frac{\sum_{i=1}^{n} X_i^4}{n\sum_{i=1}^{n} X_i^4 - \left(\sum_{i=1}^{n} X_i^2\right)^2}} = 2,086\sqrt{0,32}\sqrt{\frac{34}{5\cdot 34 - 100}} = 0,822;$$

$$B_2 = 0,95 \pm \varepsilon_2; \qquad \varepsilon_2 = t_{1-\frac{\alpha}{2}}(k)\, S\frac{1}{\sqrt{\sum_{i=1}^{n} X_i^2}} = 2,086\sqrt{0,32}\frac{1}{\sqrt{10}} = 0,373;$$

$$B_3 = -0,1774 \pm \varepsilon_3; \qquad \varepsilon_3 = t_{1-\frac{\alpha}{2}}(k)\, S \sqrt{\frac{n}{n\sum_{i=1}^{n} X_i^{\,4} - \left(\sum_{i=1}^{n} X_i^{\,2}\right)^2}} = 2,086\sqrt{0,32}\sqrt{\frac{5}{5\cdot 34 - 100}} = 0,315.$$

Полуширина доверительного интервала для коэффициента B_3 оказалась больше абсолютной величины этого коэффициента, то есть доверительный интервал для этого коэффициента накрывает значение $B_3 = 0$. В этом случае говорят, что коэффициент $\tilde{B}_3$ незначим, то есть переход от линейной модели к квадратичной не обоснован и следовало остановиться на линейной модели. Аналогичный вывод был получен при проверке адекватности полученных моделей регрессии.