

§ 9. Анализ простой линейной регрессии

Теоретические вопросы с ответами

1. Какая регрессия называется адекватной опытным данным?
2. Если имеются повторные наблюдения и на их основе вычислены дисперсия воспроизводимости D_v и дисперсия неадекватности D_n , то с помощью какого критерия судят об адекватности построенной регрессии опытным данным?
3. Если нет повторных наблюдений, то как можно судить об адекватности построенной простой линейной регрессии опытным данным по чертежу, на который нанесены экспериментальные точки и прямая регрессии y на x ?
4. Если нет повторных наблюдений, то как можно судить об адекватности построенной простой линейной регрессии опытным данным по величине выборочного коэффициента корреляции r_{XY} ?
5. Если нет повторных наблюдений, то как можно судить об адекватности построенной простой линейной регрессии опытным данным по величине остаточной дисперсии s^2 ?
6. Если нет повторных наблюдений, то по каким признакам можно судить об адекватности простой линейной регрессии опытным данным?
7. Как связаны между собой выборочный коэффициент корреляции r_{XY} и остаточная дисперсия s^2 ?
8. Если модуль выборочного коэффициента корреляции $|r_{XY}| \rightarrow 0$ то как изменяется остаточная дисперсия s^2 в случае простой линейной регрессии?
9. Если остаточная дисперсия $s^2 \rightarrow 0$, то как изменяется выборочный коэффициент корреляции r_{XY} в случае простой линейной регрессии?
10. Как можно судить об адекватности простой линейной регрессии опытным данным по двум независимым выборкам?
11. В чем состоят условия Гаусса-Маркова для экспериментальных данных?
12. Где применяются условия Гаусса-Маркова для экспериментальных данных?

13. Запишите свойство несмещенности коэффициентов B_0, B_1 эмпирической регрессии как оценок МНК соответствующих коэффициентов β_0, β_1 теоретической регрессии.
14. Запишите свойство несмещенности эмпирической регрессии $y = B_0 + B_1x$ как оценки МНК теоретической регрессии $y = \beta_0 + \beta_1x$.
15. Запишите свойство несмещенности остаточной дисперсии s^2 МНК-регрессии как оценки дисперсии σ^2 отклика Y .
16. Как сделать прогноз о значении y отклика по значению x фактора вне множества его значений?
17. Запишите уравнение прямой регрессии x на y при $r_{xy} = 0$.
18. Запишите уравнение прямой регрессии y на x при $r_{xy} = 0$.
19. Укажите угловой коэффициент $k_{y/x}$ прямой регрессии y на x .
20. Укажите угловой коэффициент $k_{x/y}$ прямой регрессии x на y .

Ответы на теоретические вопросы

1. Хорошо аппроксимирующая опытные данные.
2. Применяется критерий Фишера проверки равенства дисперсий D_e и D_n . Если они равны в пределах выбранного уровня значимости, то регрессия адекватна опытным данным.
3. Экспериментальные точки должны кучно ложиться около прямой регрессии, приблизительно в равном количестве по обе стороны от прямой.
4. Чем ближе $|r_{xy}|$ к 1, тем более кучно экспериментальные точки располагаются около прямой регрессии, и, следовательно, тем лучше эта регрессия описывает экспериментальные данные.
5. Чем меньше, тем более кучно экспериментальные точки располагаются около прямой регрессии, и, следовательно, тем лучше эта регрессия описывает экспериментальные данные.
6. Анализируются величина выборочного коэффициента корреляции r_{xy} и величина остаточной дисперсии s^2 . Чем ближе $|r_{xy}|$ к 1, а s^2 – к нулю, тем лучше аппроксимация. Также анализируется чертеж. Если экспериментальные точки располагаются близко к прямой регрессии,

приблизительно в одинаковых количествах по обе стороны от прямой, то качество аппроксимации хорошее.

$$7. s^2 = s_Y^2 \frac{n}{n-2} (1 - r_{XY}^2).$$

$$8. s^2 \rightarrow 0.$$

$$9. r_{XY} \rightarrow 1.$$

10. Строим кроме линейной регрессии другую, например, квадратичную. Сравниваем по критерию Фишера их остаточные дисперсии. Если разница между ними в пределах выбранного уровня значимости, т.е. в пределах естественного разброса, не значима, то линейная модель лучше, так как проще, а потому на данном этапе исследования адекватна опытным данным.

11. Отклонения $\varepsilon_i = y_i - y(x_i)$; ($i = 1, \dots, n$) экспериментальных ординат (величин отклика) от значений регрессии нормально распределены, взаимно независимы, имеют нулевые математические ожидания и равные дисперсии. Ошибки измерения значений фактора x_i ($i = 1, \dots, n$) пренебрежимо малы по сравнению с ошибками измерения величин y_i .

12. При построении доверительных интервалов для коэффициентов B_0, B_1 регрессии и самой регрессии $y = B_0 + B_1x$, а также при доказательстве свойства несмещенности B_0, B_1 , $y = B_0 + B_1x$ как оценок соответствующих величин теоретической регрессии.

$$13. MB_0 = \beta_0; MB_1 = \beta_1.$$

$$14. M[B_0 + B_1x] = \beta_0 + \beta_1x.$$

$$15. Ms^2 = \sigma^2$$

16. Нужно экстраполировать значения y по уравнению регрессии на один средний шаг за пределы множества значений x . Средний шаг h можно вычислить, например, по формуле $h = \frac{x_{\max} - x_{\min}}{n}$, где n – число значений фактора x .

$$17. y = \bar{y}.$$

$$18. x = \bar{x}.$$

$$19. k_{y/x} = r_{XY} \frac{s_Y}{s_X}.$$

$$20. k_{x/y} = \frac{1}{r_{XY}} \frac{s_Y}{s_X}.$$