

Тема 2.

**Парная регрессия
и корреляция**

Тема 2. Парная регрессия и корреляция

- 2.1. Основные цели и задачи регрессионного анализа
- 2.2. Постановка задачи, основные предположения регрессионного анализа
- 2.3. Парная линейная регрессия и метод наименьших квадратов
- 2.4. Меры вариации в уравнении регрессии
- 2.5. Проверка гипотез в модели парной регрессии
- 2.6. Прогнозирование в регрессионных моделях

Виды связи между явлениями (переменными Y и X):

- *Функциональная (жестко детерминированная)*. Переменные Y и X являются неслучайными, значения Y полностью определяются соответствующими значениями X , т.е. Y является некоторой функцией от переменной X (например, зависимость длины окружности от радиуса).
- *Стохастическая (случайно детерминированная)*. Зависимость Y от X проявляется в среднем (в массе случаев). В каждом отдельном случае может не проявиться в силу случайных обстоятельств. Это зависимость среднего значения Y от изменения X (например, зависимость потребления мяса от дохода):
 - *Регрессионная*. Y является случайной переменной, а X – неслучайной.
 - *Корреляционно-регрессионная*. Y и X являются случайными по своей сущности.

По направлению связи
различают:

- а) прямую;
- б) обратную.

По виду аналитической функции
различают:

- а) линейную связь;
- б) нелинейную связь.

Постановка задачи регрессии

Будем предполагать, что объясняющая переменная X оказывает воздействие на значения переменной Y , которая, таким образом, является зависимой переменной, т. е. имеет место зависимость

$$Y=f(X)$$

Постановка задачи регрессии

Пусть мы располагаем n парами выборочных наблюдений над двумя переменными X и Y : $X_1, \dots, X_n; Y_1, \dots, Y_n$

Функция $f(X)$ называется функцией регрессии Y по X , если она описывает изменение условного среднего значения результирующей переменной Y в зависимости от изменения значений объясняющей переменной X :

$$f(X) = E(Y | X).$$

Модель регрессии между Y и X имеет вид

$$Y_i = f(X_i) + \varepsilon_i,$$
$$i=1, \dots, n,$$

$f(X)$ - функция регрессии Y по X

ε – случайная составляющая (случайный член, возмущение).

Выбор вида аналитической функции $f(X)$

- используется априорная информация о содержательной экономической сущности анализируемой зависимости – аналитический способ,
- предварительный анализ зависимости с помощью визуализации – графический способ,
- использование различных статистических приемов обработки исходных данных и экспериментальных расчетов.

Парная линейная регрессия и корреляция

Пусть функция f – линейная.

Тогда модель парной линейной регрессии примет вид:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,$$

$i=1, \dots, n,$

где:

β_0 – свободный член (константа);

β_1 – коэффициент регрессии;

ε – случайная составляющая.

Показатели направления и степени тесноты связи

Для того чтобы иметь основание включить объясняющую переменную X в модель регрессии, необходимо, чтобы между переменными X и Y существовала значимая статистическая связь.

Для оценки направления и степени тесноты статистической связи используются коэффициенты ковариации, корреляции, эмпирическое и теоретическое корреляционные отношения.

Направление линейной связи можно определить с помощью линейного коэффициента ковариации.

Направление и степень тесноты линейной связи – с помощью линейного коэффициента корреляции K . Пирсона.

Коэффициент ковариации

$$\text{cov}(x, y) = \frac{\sum_{i=1}^n [(x_i - \bar{x})(y_i - \bar{y})]}{n}$$

$$\text{cov}(x, y) \in]-\infty; +\infty[$$

Для выявления влияния стажа работы (X) в годах на выработку (Y) в штуках в смену из большого количества рабочих отобраны 5 человек. Ниже приведены результаты обследования.

Рассчитать выборочные коэффициенты ковариации и корреляции. Сделать выводы.

x	y
1	2
2	4
3	8
4	6
5	10

Расчет коэффициента

ковариации

(x, y)	x	y			
	1	2	-2	-4	8
	2	4	-1	-2	2
	3	8	0	2	0
	4	6	1	0	0
	5	10	2	4	8
Сумма	15	30	0	0	18
$cov(x,y) = 3,6$					

Линейный коэффициент корреляции К.Пирсона

$$r_{x,y} = \frac{\text{cov}(x, y)}{\sigma(x)\sigma(y)}$$

$$r_{x,y} \in [-1; +1]$$

Дисперсия

$$\sigma^2(x) = \frac{\sum (x - \bar{x})^2}{n}$$

$$\sigma^2(x) = \frac{10}{5} = 2$$

Таблица 1)

	x	y					
	1	2	-2	-4	8	4	16
	2	4	-1	-2	2	1	4
	3	8	0	2	0	0	4
	4	6	1	0	0	1	0
	5	10	2	4	8	4	16
Сумма	15	30	0	0	18	10	40

Дисперсия

$$\sigma^2(y) = \frac{\sum (y - \bar{y})^2}{n}$$

$$\sigma^2(y) = \frac{40}{5} = 8$$

Среднее квадратическое отклонение

$$\sigma(x) = \sqrt{\sigma^2(x)}$$

$$\sigma(x) = \sqrt{2}$$

Среднее квадратическое отклонение

$$\sigma(y) = \sqrt{\sigma^2(y)}$$

$$\sigma(y) = \sqrt{8}$$

Линейный коэффициент корреляции К.Пирсона

$$r_{x,y} = \frac{\text{cov}(x,y)}{\sigma(x)\sigma(y)} = \frac{3,6}{\sqrt{2*8}} = 0,9$$

Коэффициент детерминации

$$r^2_{x,y} = (r_{x,y})^2 = (0,9)^2 \\ = 0,81$$

$$r^2_{x,y} \in [0; +1]$$

Коэффициент детерминации показывает, какая часть колеблемости (вариации) Y объясняется колеблемостью (вариацией) X .

Коэффициент детерминации показывает, на сколько процентов Y зависит от X .

Проверка значимости коэффициента корреляции

Формулируем гипотезы

$H_0 : \rho = 0$ (линейной корреляционной
связи между X и Y нет; коэффициент
корреляции не значим)

$H_1 : \rho \neq 0$ (между X и Y есть
линейная корреляционная связь; коэффициент
корреляции значим)

Устанавливаем уровень
значимости α

$$\alpha = 0,05$$

Находим наблюдаемое значение
критерия

$$t_{\text{набл.}} = \frac{r_{x,y}}{\sigma_r}$$

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

Находим наблюдаемое значение
критерия

$$\sigma_r = \sqrt{\frac{1 - 0,81}{5 - 2}} = 0,2517$$

$$t_{\text{набл.}} = \frac{0,9}{0,2517} = 3,58$$

Находим критическое значение критерия по таблице Стьюдента по уровню значимости α и по числу степеней свободы $k=n-m$

$$t_{\text{кр.}}(\alpha; k=n-m)$$

$$t_{\text{кр.}}(\alpha=0,05; k=5-2) = 3,18$$

t распределение: критические значения

Число степеней свободы	Двухсторонний		10%	5%	2%	1%	0.2%	0.1%
	Односторонний		5%	2.5%	1%	0.5%	0.1%	0.05%
1	6.314	12.706	31.821	63.657	318.31	636.62		
2	2.920	4.303	6.965	9.925	22.327	31.598		
3	2.353	3.182	4.541	5.841	10.214	12.924		
4	2.132	2.776	3.747	4.604	7.173	8.610		
5	2.015	2.571	3.365	4.032	5.893	6.869		
...		
...		
18	1.734	2.101	2.552	2.878	3.610	3.922		
19	1.729	2.093	2.539	2.861	3.579	3.883		
20	1.725	2.086	2.528	2.845	3.552	3.850		
...		
...		
600	1.647	1.964	2.333	2.584	3.104	3.307		
∞	1.645	1.960	2.326	2.576	3.090	3.291		

Если $|t_{набл.}| > t_{кр.}$, то нулевая гипотеза отклоняется в пользу альтернативной о статистической значимости коэффициента корреляции.

Если $|t_{набл.}| \leq t_{кр.}$, оснований отклонять нулевую гипотезу нет.

$$3,58 > 3,18$$

С надежностью, большей 0,95, и риском ошибиться, меньшим 0,05, можно утверждать, что между X и Y (между стажем и выработкой) в генеральной совокупности (для всех рабочих) существует линейная корреляционная СВЯЗЬ.

Доверительный интервал коэффициента корреляции в генеральной совокупности

- $$r_{x,y} - t_{кр.} \sigma_r < \rho < r_{x,y} + t_{кр.} \sigma_r$$

$$0,9 - 3,18 * 0,2517 < \rho < 0,9 + 3,18 * 0,2517$$

$$0,1 < \rho \leq 1$$

С надежностью 0,95 и риском ошибиться 0,05 можно утверждать, что коэффициент корреляции между X и Y (между стажем и выработкой) в генеральной совокупности (для всех рабочих) находится в интервале от 0,1 до 1.

Модель парной линейной регрессии

$$Y = \beta_0 + \beta_1 X + \varepsilon,$$

где:

β_0 - свободный член (константа);

β_1 – коэффициент регрессии;

ε – случайная составляющая.

Задачи регрессионного анализа

- Для любых значений объясняющей переменной X построить наилучшие по некоторому критерию оценки для неизвестной функции $f(X)$.
- По заданным значениям объясняющей переменной X построить наилучший по некоторому критерию прогноз для неизвестного значения результирующей переменной $Y(X)$.

Эмпирическое уравнение регрессии:

- $$\hat{y}_x = b_0 + b_1 x$$

где

b_0 и b_1 – оценки неизвестных параметров β_0 и β_1

$$y - \hat{y}_x = e$$

$$y = \hat{y}_x + e$$

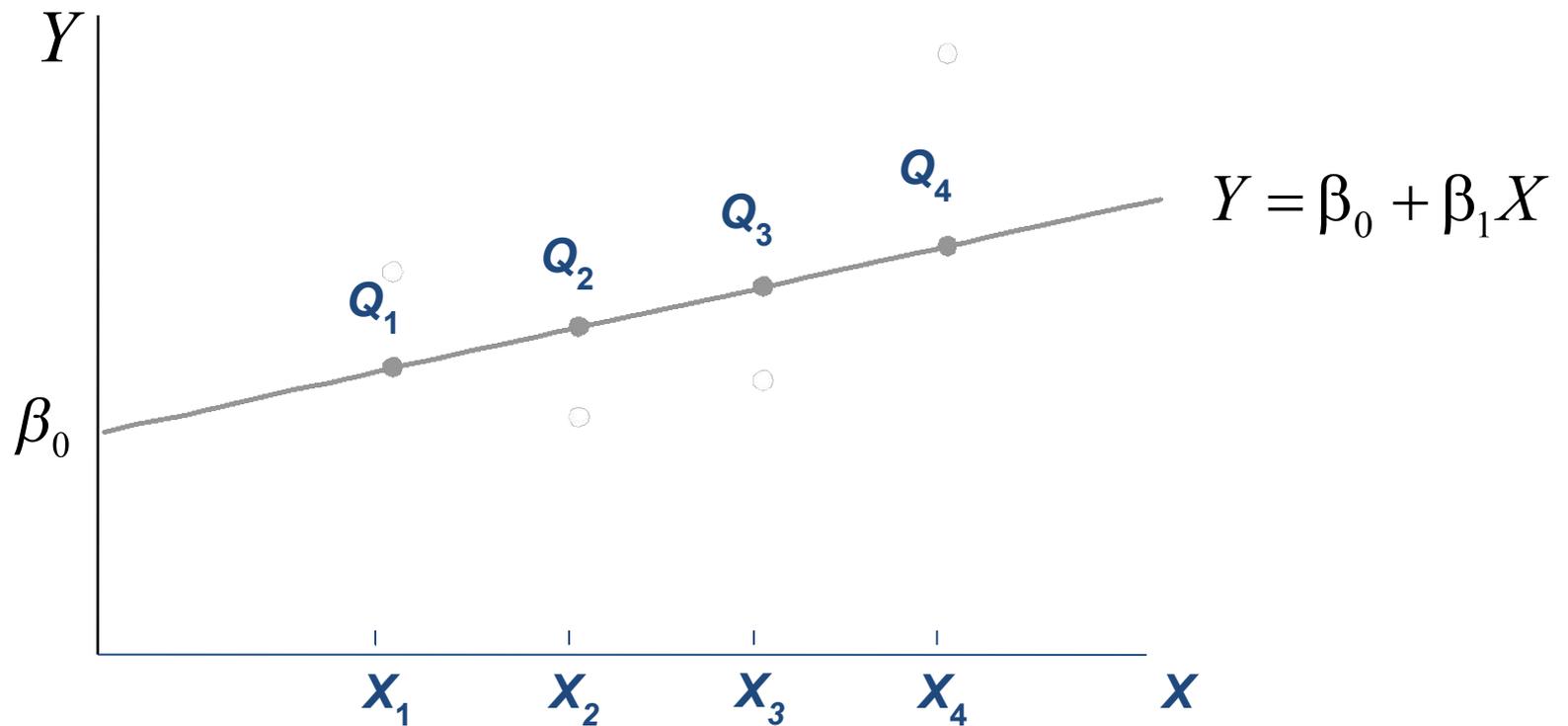
$$y = b_0 + b_1x + e$$

Модель и уравнение регрессии

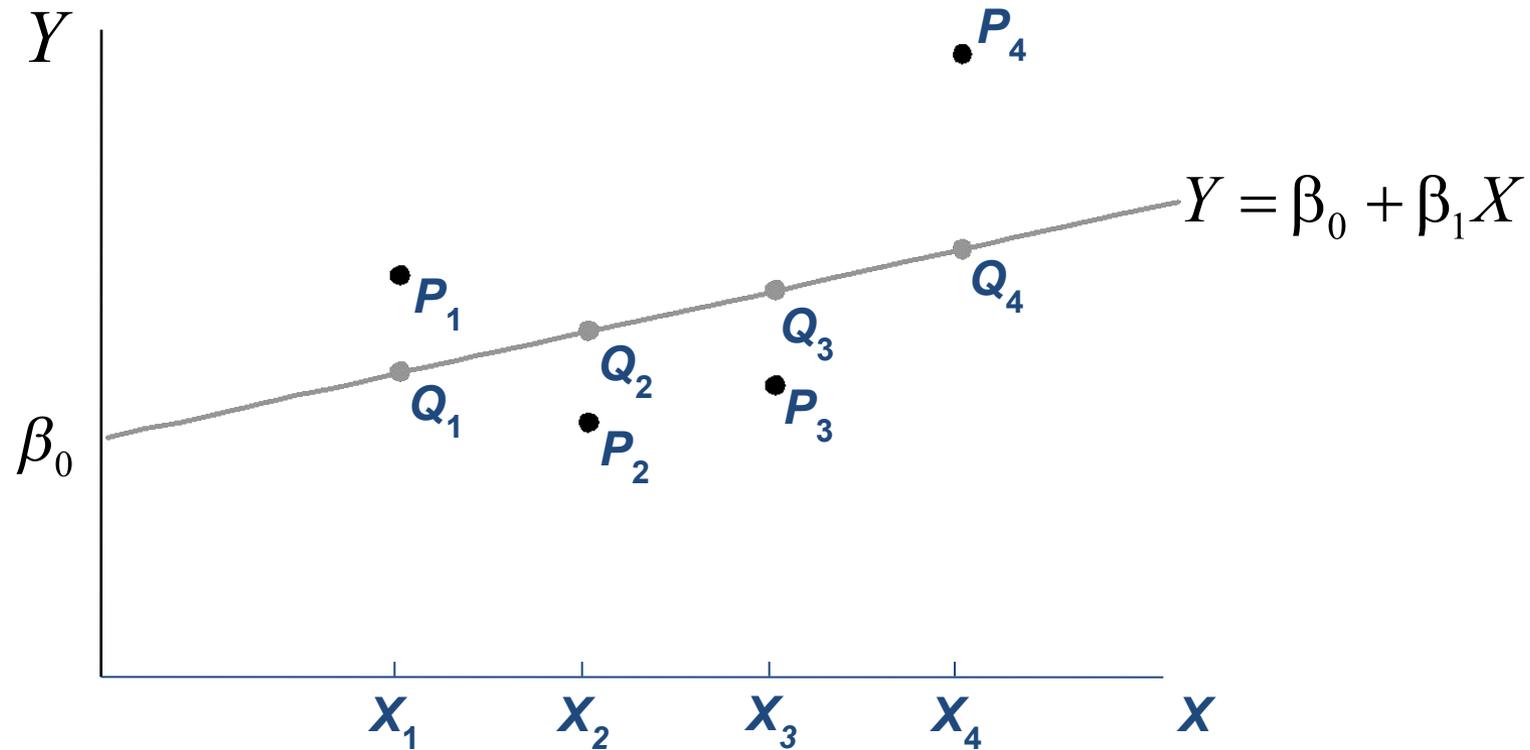
•

$$y = \beta_0 + \beta_1 x + \varepsilon$$

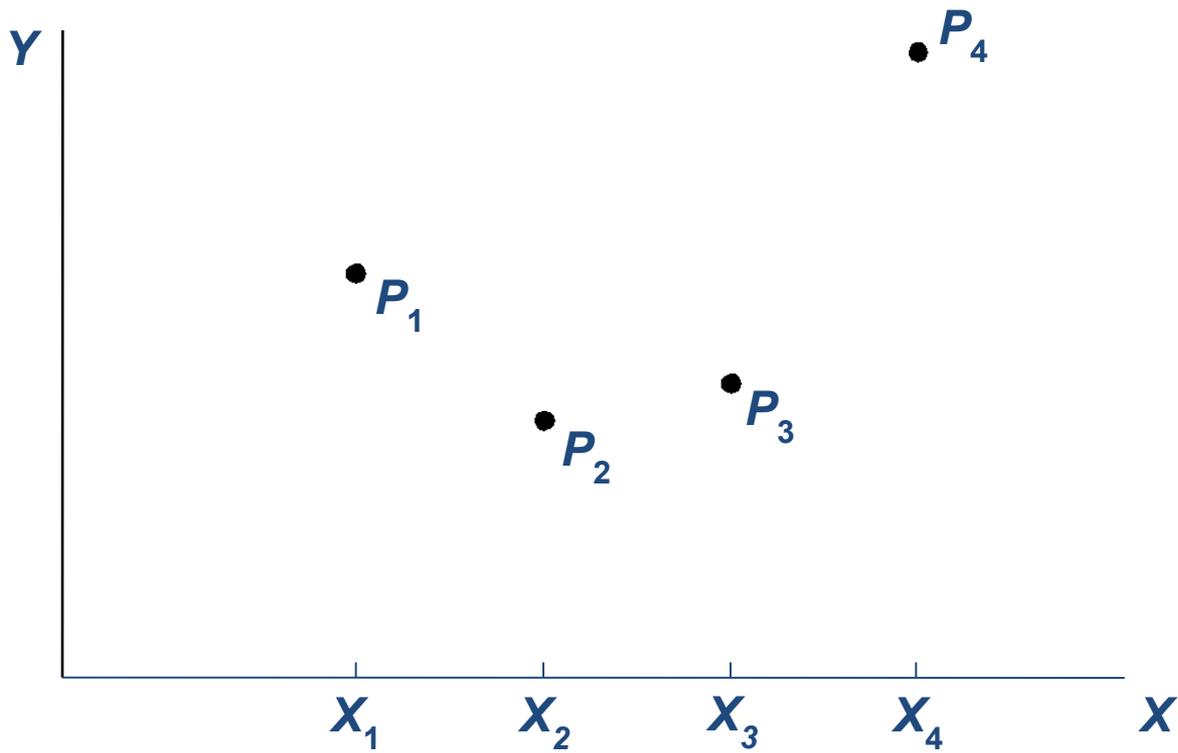
$$y = b_0 + b_1 x + e$$



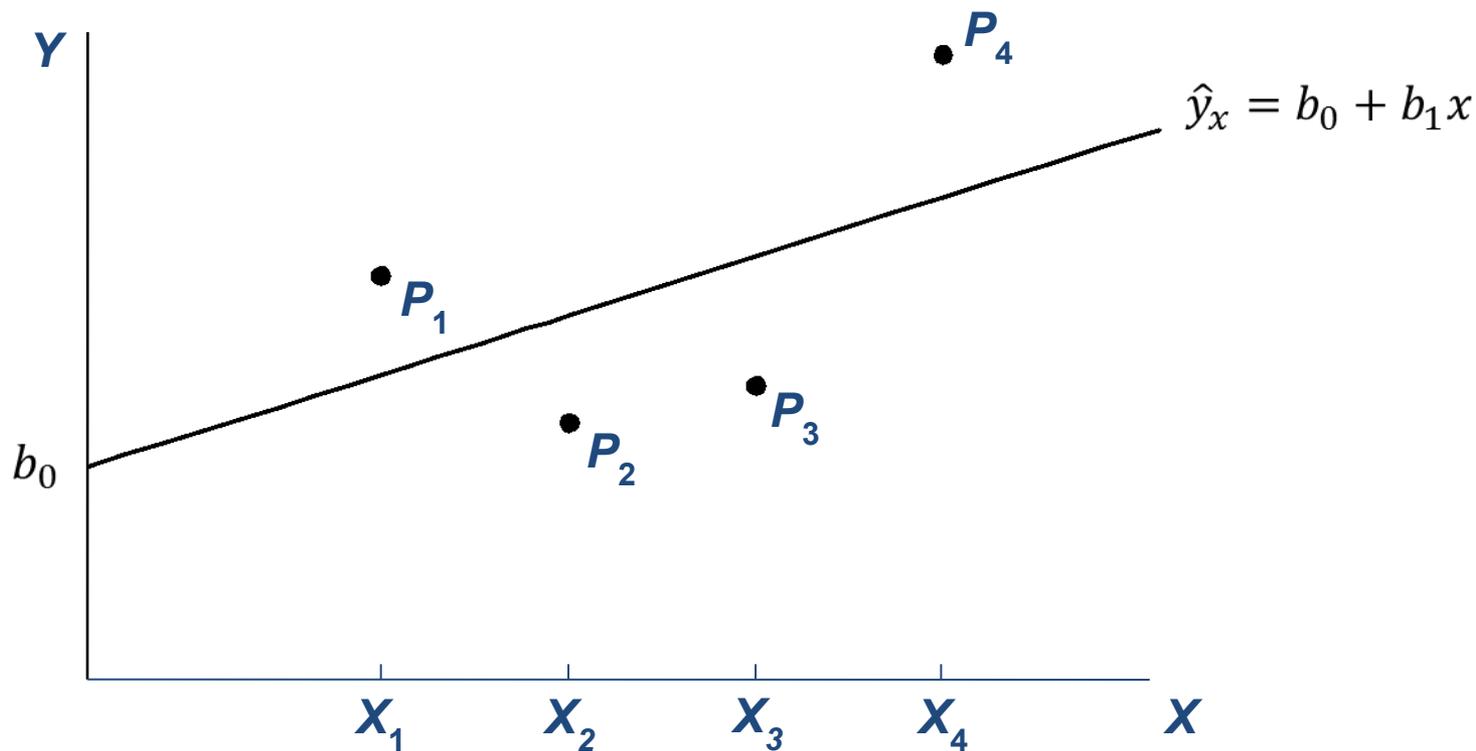
Если связь между переменными X и Y функциональная, наблюдения будут в точности лежать на прямой линии.



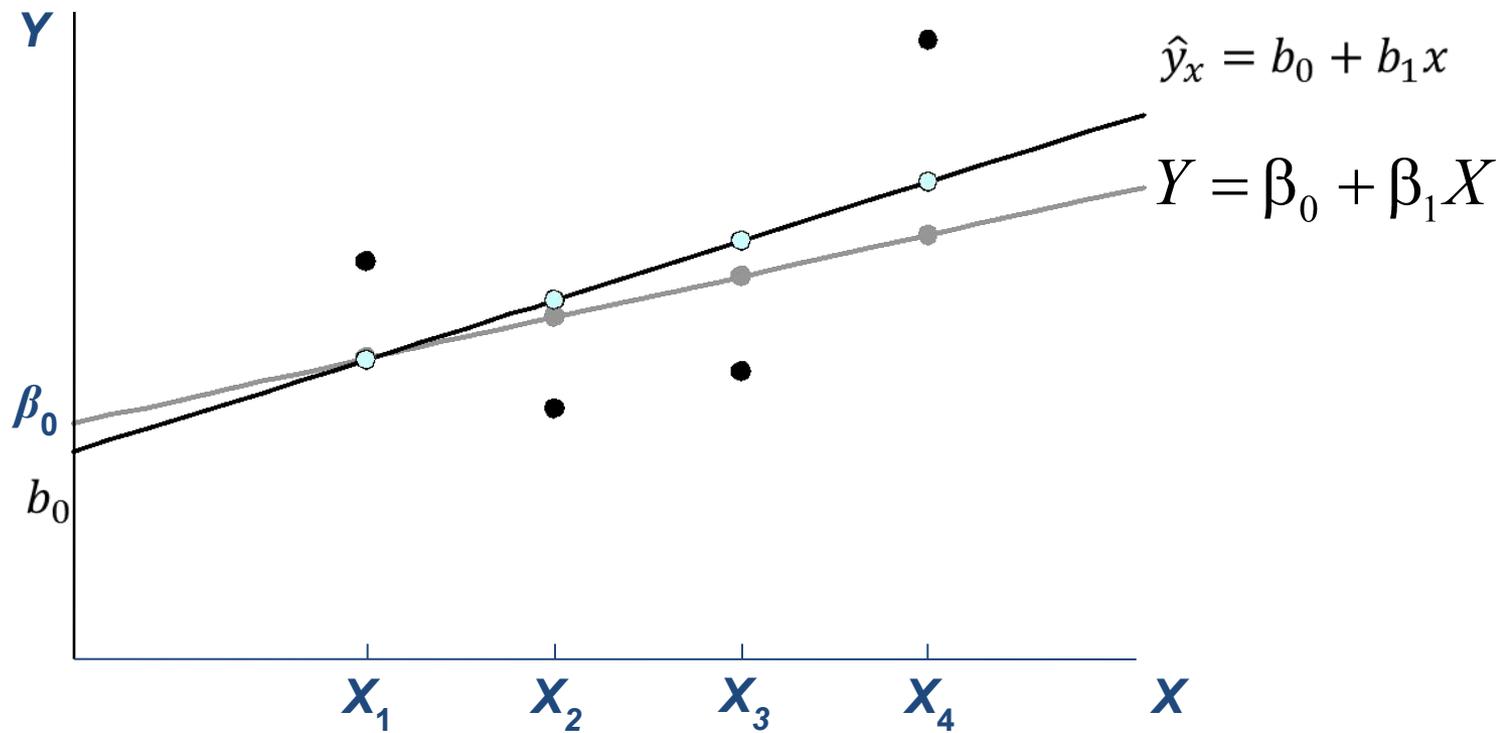
В действительности, большинство экономических связей не являются функциональными и наблюдаемые значения Y отличаются от тех, которые лежат на одной прямой.



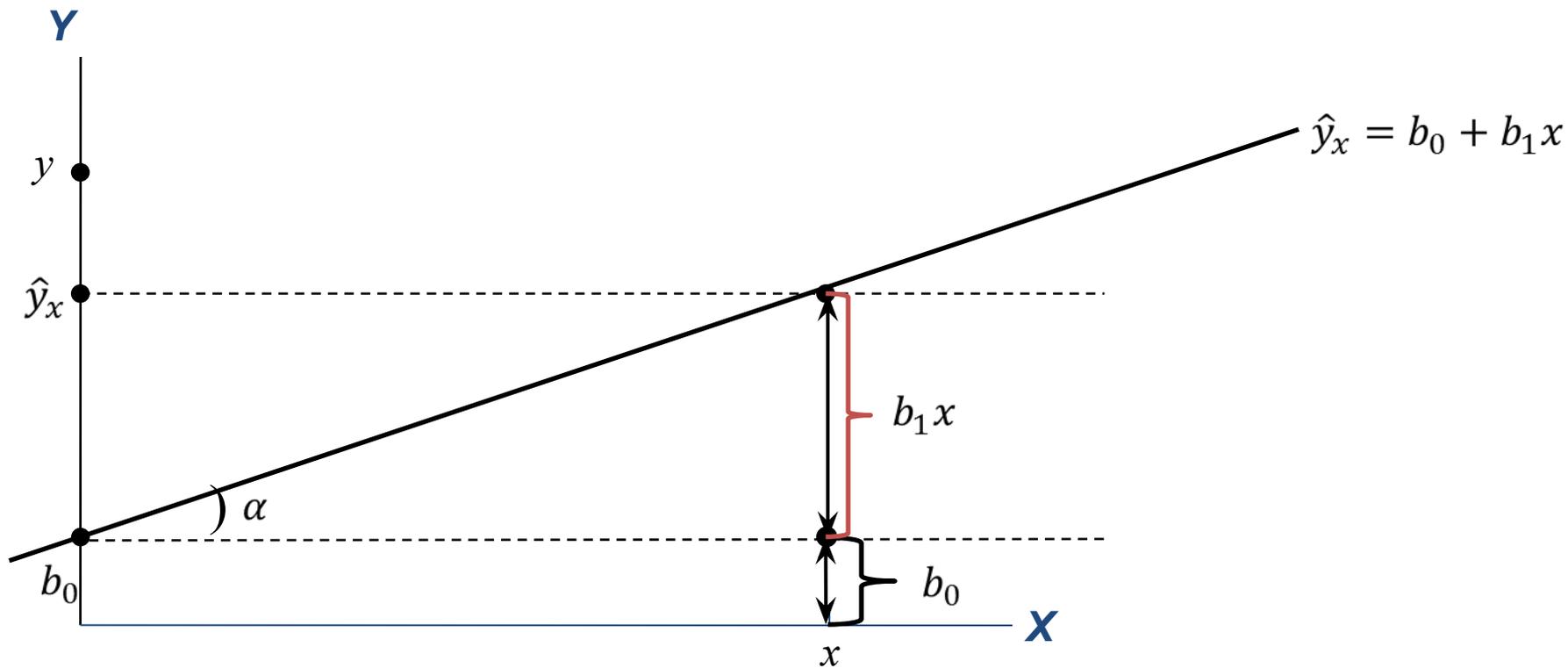
На практике мы наблюдаем только точки P .



Очевидно, мы можем использовать точки P для поиска линии, которая приближает $Y = \beta_0 + \beta_1 X + \varepsilon$. Если записать уравнение прямой $\hat{y}_x = b_0 + b_1x$ где b_0 — оценкой β_0 и b_1 — оценкой β_1 .

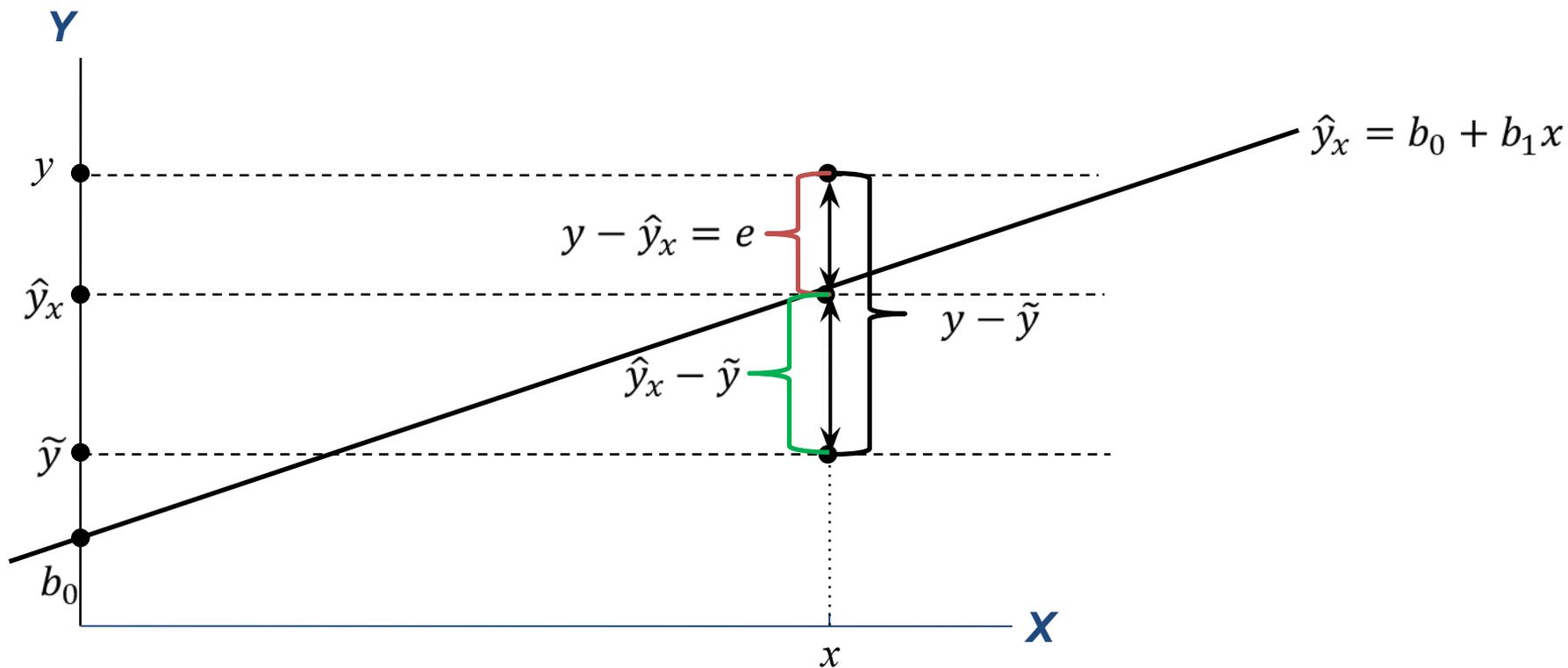


Уравнение регрессии – лишь оценка модели регрессии.



$$\operatorname{tg} \alpha = b_1$$

Метод наименьших квадратов



- $$\sum (y - \tilde{y})^2 = \sum (\hat{y}_x - \tilde{y})^2 + \sum (y - \hat{y}_x)^2$$

$$SST = SSR + SSE$$

$$SST = \sum (y - \tilde{y})^2$$

$$SSR = \sum (\hat{y}_x - \tilde{y})^2$$

$$SSE = \sum (y - \hat{y}_x)^2$$

SSR – сумма квадратов за счет регрессии (объясненная регрессией Y по X часть колеблемости Y);

SSE – сумма квадратов ошибок (остатков) (необъясненная регрессией Y по X часть колеблемости Y);

SST – общая сумма квадратов.

$$SST = SSR + SSE$$

Принцип метода наименьших квадратов

(МНК) заключается в выборе таких оценок b_0 и b_1 , для которых сумма квадратов остатков (ошибок) (e) для всех точек становится минимальной.

$$e = y - \hat{y}_x$$

Для определения оценок параметров модели регрессии b_0 и b_1 необходимо минимизировать выражение:

- $$SSE = \sum e^2 =$$
$$= \sum (y - \hat{y}_x)^2 = \sum (y - b_0 - b_1 x)^2 \rightarrow \min$$

Для этого находим частные производные первого порядка и приравниваем их к нулю:

$$\begin{cases} \frac{dSSE}{db_0} = -2 \sum (y - b_0 - b_1 x)^2 = 0 \\ \frac{dSSE}{db_1} = -2 \sum x(y - b_0 - b_1 x)^2 = 0 \end{cases}$$

Отсюда получим формулы расчета оценок параметров модели регрессии

$$b_1 = \frac{\text{cov}(x, y)}{\sigma^2(x)}$$

$$b_0 = \tilde{y} - b_1 \tilde{x}$$

Расчет оценок параметров модели регрессии

$$b_1 = \frac{3,6}{2} = 1,8$$

$$b_0 = 6 - 1,8 * 3 = 0,6$$

Уравнение регрессии

$$\hat{y}_x = b_0 + b_1x$$

$$\hat{y}_x = 0,6 + 1,8x$$

Интерпретация коэффициента регрессии

Коэффициент регрессии b_1 показывает на сколько единиц увеличится (уменьшится) значение зависимой переменной Y (в единицах измерения переменной Y) при увеличении (уменьшении) значения объясняющей переменной X на одну единицу (в единицах измерения переменной X).

Интерпретация свободного члена

Свободный член b_0 показывает базисный (начальный) уровень, т.е. значение зависимой переменной Y при условии, что объясняющая переменная X равна нулю.

В случае, если такая интерпретация лишена экономического смысла, свободный член интерпретируется как параметр, отражающий агрегированное влияние переменных, не включенных в модель.

Интерпретация коэффициента регрессии

Коэффициент регрессии b_1 показывает, что при увеличении стажа на 1 год выработка в среднем увеличится на 1,8 штуки в смену.

Интерпретация свободного члена

Свободный член b_0 показывает, что выработка рабочего, не имеющего стажа, составит 0,6 штуки в смену.

Проверка статистической значимости уравнения регрессии в целом.

Сформулируем гипотезы:

$$H_0 : R^2 = 0$$

Y не зависит от всех X, включенных в модель (уравнение в целом не значимо)

$$H_1 : R^2 > 0$$

Y зависит от всех X (вместе взятых), включенных в модель (уравнение в целом значимо)

Устанавливаем уровень
значимости α

$$\alpha = 0,05$$

Найдем наблюдаемое значение критерия

$$F_{\text{набл.}} = \frac{SSR / (m - 1)}{SSE / (n - m)}$$

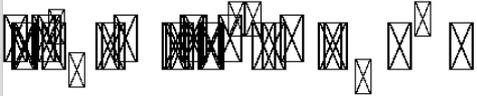
где n – число наблюдений,

m – число параметров в модели
регрессии (для парной регрессии $m=2$)

Расчет SSR, SSE и SST

	x	y				
	1	2	2,4	-0,4	-3,6	-4
	2	4	4,2	-0,2	-1,8	-2
	3	8	6	2	0	2
	4	6	7,8	-1,8	1,8	0
	5	10	9,6	0,4	3,6	4
Сумма	15	30	30	0	0	0

Расчет SSR, SSE и SST

		
0,16	12,96	16
0,04	3,24	4
4	0	4
3,24	3,24	0
0,16	12,96	16
7,6	32,4	40
SSE	SSR	SST

Найдем наблюдаемое значение критерия

$$F_{\text{набл.}} = \frac{32,4 / (2 - 1)}{7,6 / (5 - 2)} = 12,78$$

По таблице распределения Фишера найдем критическое значение критерия:

$$F_{кр} = F(\alpha; m - 1; n - m)$$

$$F_{кр} = F(\alpha = 0,05; 2 - 1; 5 - 2) = 10,13$$

Уровень значимости $\alpha = 0,05$

$K_2 \backslash K_1$	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38

Если $F_{набл.} > F_{кр.}$, то нулевая гипотеза отклоняется в пользу альтернативной о статистической значимости уравнения регрессии в целом. Если $F_{набл.} \leq F_{кр.}$, оснований отклонять нулевую гипотезу нет.

$$12,78 > 10,13$$

С надежностью, большей 0,95, и риском ошибиться, меньшим 0,05, можно утверждать, что Y (выработка) зависит от всех X , включенных в модель (от стажа).

Проверка статистической значимости
коэффициента регрессии
Сформулируем гипотезы

$$H_0 : \beta_1 = 0$$

Y не зависит от
данного конкретного
X (коэффициент
регрессии не значим)

$$H_1 : \beta_1 \neq 0$$

Y зависит от данного
конкретного X
(коэффициент
регрессии значим)

Устанавливаем уровень
значимости α

$$\alpha = 0,05$$

Находим наблюдаемое значение критерия

$$t_{\text{набл.}} = \frac{b_1}{S_{b_1}}$$

S_{b_1} - стандартная ошибка коэффициента регрессии

$$S_{b_1} = \sqrt{\frac{S^2_{yx}}{n\sigma^2(x)}}$$

Стандартная ошибка уравнения регрессии

$$S_{yx} = \sqrt{\frac{SSE}{n - m}}$$

- $$S_{yx} = \sqrt{\frac{7,6}{5 - 2}} = \sqrt{2,53} = 1,5916$$

Стандартная ошибка коэффициента регрессии

$$S_{b_1} = \sqrt{\frac{2,53}{5 * 2}} = 0,5033$$

Находим наблюдаемое значение
критерия

$$t_{\text{набл.}} = \frac{1,8}{0,5033} = 3,58$$

Находим критическое значение критерия по таблице Стьюдента по уровню значимости α и по числу степеней свободы $k=n-m$

$$t_{\text{кр.}}(\alpha; k=n-m)$$

$$t_{\text{кр.}}(\alpha=0,05; k=5-2) = 3,18$$

t распределение: критические значения

Число степеней свободы	Двухсторонний		10%	5%	2%	1%	0.2%	0.1%
	Односторонний		5%	2.5%	1%	0.5%	0.1%	0.05%
1	6.314	12.706	31.821	63.657	318.31	636.62		
2	2.920	4.303	6.965	9.925	22.327	31.598		
3	2.353	3.182	4.541	5.841	10.214	12.924		
4	2.132	2.776	3.747	4.604	7.173	8.610		
5	2.015	2.571	3.365	4.032	5.893	6.869		
8	...	2,306			
...				
18	1.734	2.101	2.552	2.878	3.610	3.922		
19	1.729	2.093	2.539	2.861	3.579	3.883		
20	1.725	2.086	2.528	2.845	3.552	3.850		
...				
600	1.647	1.964	2.333	2.584	3.104	3.307		
∞	1.645	1.960	2.326	2.576	3.090	3.291		

Если $|t_{набл.}| > t_{кр.}$, то нулевая гипотеза отклоняется в пользу альтернативной о статистической значимости коэффициента регрессии.

Если $|t_{набл.}| \leq t_{кр.}$, оснований отклонять нулевую гипотезу нет.

$$3,58 > 3,18$$

С надежностью, большей 0,95, и риском ошибиться, меньшим 0,05, можно утверждать, что Y (выработка) зависит от данного конкретного X (от стажа).

Проверка статистической значимости свободного члена

Сформулируем гипотезы

$$H_0 : \beta_0 = 0$$

Свободный член не
значим (незначимо
отличается от 0)

$$H_1 : \beta_0 \neq 0$$

Свободный член
значим (значимо
отличается от 0)

Наблюдаемое значение критерия

$$t_{\text{набл}} = \frac{b_0}{S_{b_0}}$$

S_{b_0} - стандартная ошибка
свободного члена

Стандартная ошибка свободного члена:

$$S_{b_0} = \sqrt{\frac{S^2_{yx}}{n} \left(1 + \frac{\tilde{x}^2}{\sigma^2(x)}\right)}$$

$$S_{b_0} = \sqrt{\frac{2,53}{5} \left(1 + \frac{3^2}{2}\right)} = 1,6693$$

Наблюдаемое значение критерия

$$t_{\text{набл}} = \frac{0,6}{1,6693} = 0,36$$

Находим критическое значение критерия по таблице Стьюдента по уровню значимости α и по числу степеней свободы $k=n-m$

$$t_{\text{кр.}}(\alpha; k=n-m)$$

$$t_{\text{кр.}}(\alpha=0,05; k=5-2) = 3,18$$

Если $|t_{набл.}| > t_{кр.}$, то нулевая гипотеза отклоняется в пользу альтернативной о статистической значимости свободного члена.

Если $|t_{набл.}| \leq t_{кр.}$, оснований отклонять нулевую гипотезу нет.

$$0,36 < 3,18$$

На уровне значимости $\alpha=0,05$
свободный член не значим.

Доверительные интервалы неизвестных значений β_1 и β_0

$$b_1 - t_{\text{кр}} S_{b_1} < \beta_1 < b_1 + t_{\text{кр}} S_{b_1}$$

$$b_0 - t_{\text{кр}} S_{b_0} < \beta_0 < b_0 + t_{\text{кр}} S_{b_0}$$

Доверительный интервал неизвестного значения β_1

- $$b_1 - t_{\text{кр.}} S_{b_1} < \beta_1 < b_1 + t_{\text{кр.}} S_{b_1}$$

$$1,8 - 3,18 * 0,5033 < \beta_1 < 1,8 + 3,18 * 0,5033$$

$$0,2 < \beta_1 < 3,4$$

С надежностью 0,95 и риском ошибиться 0,05 можно утверждать, что коэффициент регрессии в генеральной совокупности (для всех рабочих) находится в интервале от 0,2 до 3,4.

При увеличении стажа на 1 год выработка в среднем увеличится от 0,2 до 3,4 штуки в смену.

Так как интервал не включает 0, коэффициент регрессии значим.

Доверительный интервал неизвестного значения β_0

- $$b_0 - t_{\text{кр.}} S_{b_0} < \beta_0 < b_0 + t_{\text{кр.}} S_{b_0}$$

$$0,6 - 3,18 * 1,6693 < \beta_0 < 0,6 + 3,18 * 1,6693$$

$$-4,71 < \beta_0 < 5,91$$

С надежностью 0,95 и риском ошибиться 0,05 можно утверждать, что свободный член в генеральной совокупности (для всех рабочих) находится в интервале от -4,71 до 5,91.

Так как интервал включает 0, свободный член не значим.

Точечный прогноз по уравнению
регрессии

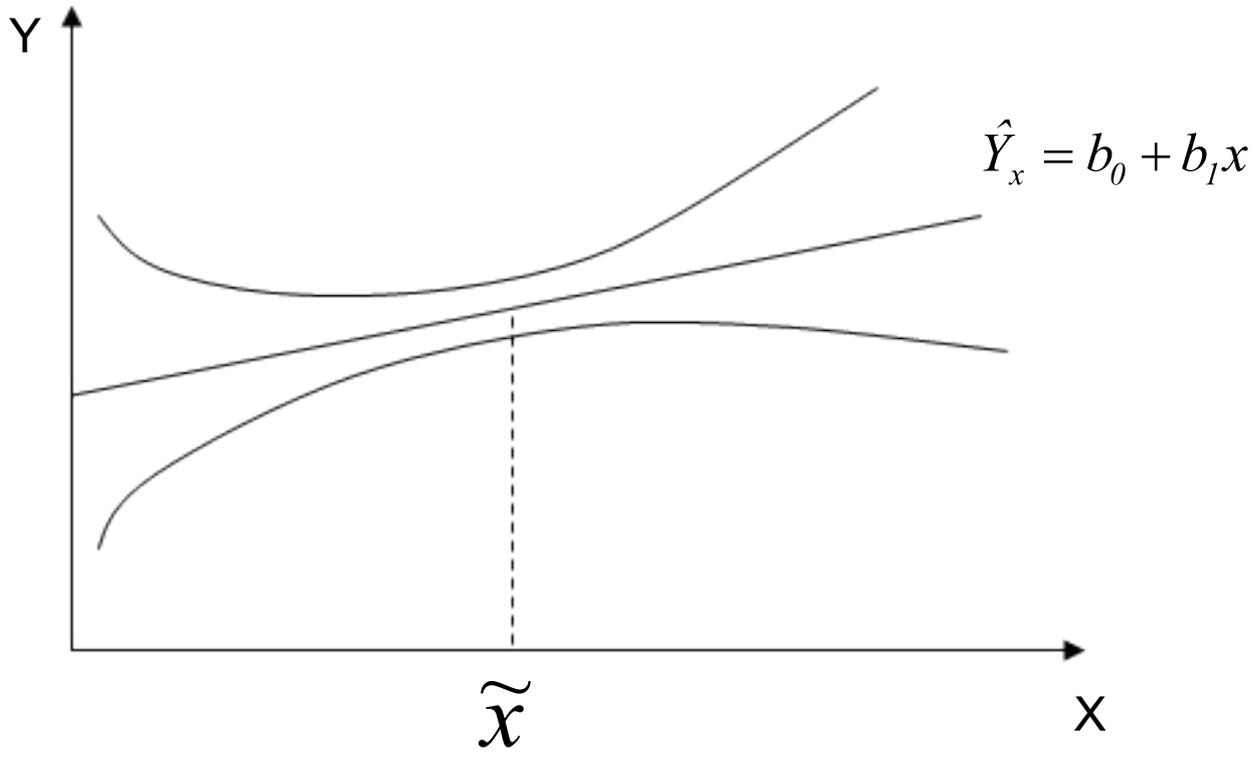
$$\hat{y}_{x^*} = b_0 + b_1 x^*$$

Точечный прогноз по уравнению
регрессии

$$\hat{y}_x = 0,6 + 1,8x$$

$$x^* = 2,5$$

$$\hat{y}_{x=2,5} = 0,6 + 1,8 * 2,5 = 5,1$$



Интервальный прогноз неизвестного среднего генерального значения Y

$$\hat{Y}_{x^*} - t_{кр} S_{yx} \sqrt{h^*} < \bar{Y}_z < \hat{Y}_{x^*} + t_{кр} S_{yx} \sqrt{h^*}$$

$$h^* = \frac{1}{n} + \frac{(x^* - \tilde{x})^2}{\sum_{i=1}^n (x_i - \tilde{x})^2} = \frac{1}{n} + \frac{(x^* - \tilde{x})^2}{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}}$$

$$t_{kp} = 3,18$$

$$S_{yx} = 1,5916$$

$$h^* = \frac{1}{5} + \frac{(2,5 - 3)^2}{55 - \frac{15^2}{5}} = 0,225$$

Интервальный прогноз неизвестного среднего генерального значения \bar{Y}

$$5,1 - 3,18 * 1,5916 * \sqrt{0,225} < \bar{Y}_2 < 5,1 + 3,18 * 1,5916 * \sqrt{0,225}$$

$$2,7 < \bar{Y}_2 < 7,5$$

Интервальный прогноз неизвестного индивидуального значения Y

$$\hat{Y}_{x^*} - t_{кр} S_{yx} \sqrt{1 + h^*} < Y_u < \hat{Y}_{x^*} + t_{кр} S_{yx} \sqrt{1 + h^*}$$

Интервальный прогноз неизвестного индивидуального значения Y

$$5,1 - 3,18 * 1,5916 * \sqrt{1 + 0,225} < Y_u < 5,1 + 3,18 * 1,5916 * \sqrt{1 + 0,225}$$

$$5,1 - 5,6 < Y_u < 5,1 + 5,6$$

$$0 < Y_u < 10,7$$