ТИПЫ РЕГРЕССИОННОЙ МОДЕЛИ И ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

ТИПЫ ДАННЫХ

Перекрестные: Наблюдения за домашними хозяйствами, предприятиями, странами и т. д. в один момент времени (модели A и B).

Временные ряды: Наблюдения за доходами, потреблением, процентными ставками и т. д. в течение ряда периодов времени (модель C).

Временные ряды: Наблюдения за индивидами, домашними хозяйствами и т. д. в течение ряда периодов времени (Глава 14, модель B).

В течение данного курса лекций мы будем работать с тремя типами данных, описанных выше.

ТИПЫ РЕГРЕССИОННОЙ МОДЕЛИ И ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

ТИПЫ МОДЕЛЕЙ

Модель A:	**Перекрестные данные с нестохастическими независимыми переменными в регрессии. Их значения в наблюдениях в объеме выборки не содержат случайных (хаотичных) компонентов.**
Модель B:	**Перекрестные данные со стохастическими независимыми переменными в регрессии. Значения независимых переменных регрессии описаны случайным образом и независимо от заданной генеральной совокупности.**
Модель C:	**Данные временного ряда. Независимые переменные в регрессии могут сохранять свои значения в течение продолжительного периода времени. Регрессии с данными временного ряда потенциально включают совокупность технических проблем, которые наилучшим образом предотвращаются изначально.**

Разные регрессионные модели подходят для различных типов данных. Мы будем рассматривать три типа регрессионной модели, приведенных выше.

ТИПЫ РЕГРЕССИОННОЙ МОДЕЛИ И ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

ТИПЫ МОДЕЛЕЙ

Модель А:	**Перекрестные данные с нестохастическими независимыми переменными в регрессии. Их значения в наблюдениях в объеме выборки не содержат случайных (хаотичных) компонентов.**
Модель B:	**Перекрестные данные со стохастическими независимыми переменными в регрессии. Значения независимых переменных регрессии описаны случайным образом и независимо от заданной генеральной совокупности.**
Модель C:	**Данные временного ряда. Независимые переменные в регрессии могут сохранять свои значения в течение продолжительного периода времени. Регрессии с данными временного ряда потенциально включают совокупность технических проблем, которые наилучшим образом предотвращаются изначально.**

Начнем с модели А. Будем делать так исключительно для аналитического удобства. Это позволит нам проводить исследование регрессионного анализа в относительно простых рамках классической линейной регрессионной модели.

ТИПЫ РЕГРЕССИОННОЙ МОДЕЛИ И ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

ТИПЫ МОДЕЛЕЙ

Модель A: Перекрестные данные с нестохастическими независимыми переменными в регрессии. Их значения в наблюдениях в объеме выборки не содержат случайных (хаотичных) компонентов.

Модель B: Перекрестные данные со стохастическими независимыми переменными в регрессии. Значения независимых переменных регрессии описаны случайным образом и независимо от заданной генеральной совокупности.

Модель C: Данные временного ряда. Независимые переменные в регрессии могут сохранять свои значения в течение продолжительного периода времени. Регрессии с данными временного ряда потенциально включают совокупность технических проблем, которые наилучшим образом предотвращаются изначально.

Заменим это в Главе 8 более простым и реалистичным предположением, подходящим для регрессий с перекрестными данными таким образом, что наблюдения за независимыми переменными регрессии выводятся из заданной генеральной совокупности.

ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

А.1 Модель линейна относительно параметров и точно определена.

Например:

$$Y = \beta_1 + \beta_2 X + u$$

Примеры моделей, которые являются нелинейными относительно параметров:

$$Y = \beta_1 X^{\beta_2} + u$$

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_2 \beta_3 X_4 + u$$

‘Линейный относительно параметров’ означает, что каждый член с правой стороны включает β как простой фактор и среди β нет встроенной связи. Отложим обсуждение вопросов, касающихся линейности и нелинейности, к Главе 4.

ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

A.2 В выборке есть коэффициент вариации независимой переменной регрессии.

В выборке должна быть вариация независимой переменной регрессии. В противном случае не будет объяснено любое изменение *Y*.

ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

А.2 В выборке есть коэффициент вариации независимой переменной регрессии.

$$b_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

Если $X_i = \bar{X}$ для всех *i*,

$$b_2 = \frac{0}{0}$$

Если мы попытаемся оценить влияние *Y на X,* когда значения *X* постоянны, мы обнаружим, что мы не сможем вычислить коэффициенты регрессии. Как числитель, так и знаменатель выражения для b_2 были бы равны нулю. Мы также не сможем получить значения b_1.

ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

A.3 Остаточный член имеет нулевое ожидание

$$E(u_i) = 0 \quad \text{для всех } i$$

Предположим, что ожидаемое значение остаточного члена при любом наблюдении должно быть равным нулю. В некоторых случаях остаточный член будет положительным, а в некоторых отрицательным, но он не должен иметь тенденцию в любом направлении.

ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

А.3 Остаточный член имеет нулевое ожидание

$$E(u_i) = 0 \quad \text{для всех } i$$

Фактически, если свободный член включен в уравнение регрессии, разумно предположить, что это условие выполняется автоматически. Роль свободного члена заключается в отборе некоторой систематической, но постоянной тенденции *Y*, необъясняемой независимой (-ыми) переменной (-ыми) регрессии.

ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

А.3 Остаточный член имеет нулевое ожидание

$$E(u_i) = 0$$ **для всех *i***

Предположим $Y = \beta_1 + \beta_2 X + u$ $\qquad E(u_i) = \mu_u \neq 0$

Предположим, что свободный член имеет ненулевое генеральное среднее.

ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

А.3 Остаточный член имеет нулевое ожидание

$$E(u_i) = 0$$ для всех *i*

Предположим $Y = \beta_1 + \beta_2 X + u$ $E(u_i) = \mu_u \neq 0$

Введем $v_i = u_i - \mu_u$

Введем новую случайную переменную $v_i = u_i - \mu_u$.

ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ A

A.3 Остаточный член имеет нулевое ожидание

$$E(u_i) = 0 \quad \text{для всех } i$$

Предположим $Y = \beta_1 + \beta_2 X + u \qquad E(u_i) = \mu_u \neq 0$

Введем $v_i = u_i - \mu_u$

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_i + v_i + \mu_u \\ &= \beta_1^* + \beta_2 X_i + v_i \qquad \text{где} \qquad \beta_1^* = \beta_1 + \mu_u \end{aligned}$$

Затем можем перезаписать модель, как показано выше. v_i становится новым остаточным членом и свободный член включает постоянную μ_u.

ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

А.3 Остаточный член имеет нулевое ожидание

$E(u_i) = 0$ для всех i

Предположим $Y = \beta_1 + \beta_2 X + u$ $E(u_i) = \mu_u \neq 0$

Введем $v_i = u_i - \mu_u$

$$Y_i = \beta_1 + \beta_2 X_i + v_i + \mu_u$$

$$= \beta_1^* + \beta_2 X_i + v_i$$ где $\beta_1^* = \beta_1 + \mu_u$

Тогда $E(v_i) = E(u_i - \mu_u) = E(u_i) - E(\mu_u) = \mu_u - \mu_u = 0$

Остаточный член в скорректированной модели теперь соответствует предположению А.3.

ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

A.3 Остаточный член имеет нулевое ожидание

$$E(u_i) = 0 \quad \text{для всех } i$$

Предположим $Y = \beta_1 + \beta_2 X + u \qquad E(u_i) = \mu_u \neq 0$

Введем $v_i = u_i - \mu_u$

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_i + v_i + \mu_u \\ &= \beta_1^* + \beta_2 X_i + v_i \qquad \text{где} \qquad \beta_1^* = \beta_1 + \mu_u \end{aligned}$$

Тогда $E(v_i) = E(u_i - \mu_u) = E(u_i) - E(\mu_u) = \mu_u - \mu_u = 0$

Ценность заключается в том, что интерпретация свободного члена изменилась. Он поглотил ненулевую составляющую остаточного члена.

ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

А.3 Остаточный член имеет нулевое ожидание

$E(u_i) = 0$ для всех *i*

Предположим $Y = \beta_1 + \beta_2 X + u$ $\quad E(u_i) = \mu_u \neq 0$

Введем $v_i = u_i - \mu_u$

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_i + v_i + \mu_u \\ &= \beta_1^* + \beta_2 X_i + v_i \end{aligned}$$

где $\beta_1^* = \beta_1 + \mu_u$

Тогда $E(v_i) = E(u_i - \mu_u) = E(u_i) - E(\mu_u) = \mu_u - \mu_u = 0$

Это допустимо, потому что роль постоянной обычно заключается в отборе некоторой систематической тенденции *Y*, необъясняемой независимой (-ыми) переменной (-ыми) регрессии.

ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

A.4 Остаточный член гомоскедастичный (с постоянной условной дисперсией)

$$\sigma_{u_i}^2 = \sigma_u^2 \quad \text{для всех } i$$

Предположим, что остаточный член гомоскедастичен, что означает, что его значение в каждом наблюдении получается из распределения с постоянной дисперсией генеральной совокупности.

ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

A.4 Остаточный член гомоскедастичный (с постоянной условной дисперсией)

$$\sigma_{u_i}^2 = \sigma_u^2 \qquad \text{для всех } i$$

На языке раздела по выборочному наблюдению и метода оценивания в главе «Пересмотр» это предварительная концепция, в которой мы думаем о возможном поведении остаточного члена до того, как фактически будет сгенерирована выборка.

ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

A.4 Остаточный член гомоскедастичный (с постоянной условной дисперсией)

$$\sigma_{u_i}^2 = \sigma_u^2 \qquad \text{для всех } i$$

После того, как мы сгенерировали выборку, остаточный член в некоторых наблюдениях окажется больше, а в некоторых меньше, но не должно быть никакой причины, по которой он будет в некоторых наблюдениях более неустойчивым, чем в других.

ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

А.4 Остаточный член гомоскедастичный (с постоянной условной дисперсией)

$$\sigma_{u_i}^2 = \sigma_u^2 \quad \text{для всех } i$$

$$\sigma_{u_i}^2 = E\left((u_i - \mu_u)^2\right) = E\left(u_i^2\right)$$

$$E\left(u_i^2\right) = \sigma_u^2 \quad \text{для всех } i$$

Так как $E(u_i) = 0$, то согласно предположению А.3, дисперсия генеральной совокупности u_i равна $E(u_i^2)$, поэтому условие может быть также записано, как показано выше.

ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

А.4 Остаточный член гомоскедастичный (с постоянной условной дисперсией)

$$\sigma_{u_i}^2 = \sigma_u^2 \quad \text{для всех } i$$

$$\sigma_{u_i}^2 = E\left((u_i - \mu_u)^2\right) = E\left(u_i^2\right)$$

$$E\left(u_i^2\right) = \sigma_u^2 \quad \text{для всех } i$$

Если предположение А.4 не выполняется, коэффициенты регрессии МНК будут неэффективными, и вы сможете получить более надежные результаты, используя модификацию метода регрессии. Это будет рассмотрено в Главе 7.

ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

A.5 Значения остаточного члена имеют независимые распределения

u_i распределены независимо от u_j для всех $j \neq i$

Мы предполагаем, что остаточный член не подлежит автокорреляции, а значит, не должно быть систематической связи между его значениями в любых двух наблюдениях.

ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

A.5 Значения остаточного члена имеют независимые распределения

u_i распределены независимо от u_j для всех $j \neq i$

Например, только потому, что остаточный член является большим и положительным в одном наблюдении, не должно быть тенденции, чтобы он был большим и положительным в следующем (или большим и отрицательным, а также, или малым и положительным, или малым и отрицательным).

ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

A.5 Значения остаточного члена имеют независимые распределения

u_i распределены независимо от u_j для всех $j \neq i$

$$\sigma_{u_i u_j} = E\left[(u_i - \mu_u)(u_j - \mu_u)\right] = E(u_i u_j)$$
$$= E(u_i)E(u_j) = 0$$

Из предположения следует, что ковариация генеральной совокупности между u_i и u_j равна нулю. Заметим, что генеральное среднее u_i и u_j равны нулю в силу предположения A.3, и что E(u_iu_j) можно разложить как E(u_i)E(u_j), если u_i and u_j генерируются независимо – смотреть «Пересмотр».

ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

A.5 Значения остаточного члена имеют независимые распределения

u_i распределены независимо от u_j для всех $j \neq i$

$$\sigma_{u_i u_j} = E\left[(u_i - \mu_u)(u_j - \mu_u)\right] = E(u_i u_j)$$
$$= E(u_i)E(u_j) = 0$$

Если это предположение не будет выполнено, МНК снова даст неэффективные оценки. В Главе 12 рассмотрены возникающие проблемы и способы их избежания. Несоблюдения этого предположения редко встречаются с перекрестными данными.

ТИПЫ РЕГРЕССИОННОЙ МОДЕЛИ И ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

A.6 Остаточный член имеет нормальное распределение

Обычно мы предполагаем, что остаточный член имеет нормальное распределение. Обоснование предположения основано на теореме Центрального предела.

ТИПЫ РЕГРЕССИОННОЙ МОДЕЛИ И ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

А.6 Остаточный член имеет нормальное распределение

В сущности, центральная предельная теорема утверждает, что если случайная переменная является результатом влияния огромного количества других случайных переменных, она будет иметь приблизительно нормальное распределение, даже если все ее компоненты не будут доминирующими.

ТИПЫ РЕГРЕССИОННОЙ МОДЕЛИ И ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

ПРЕДПОЛОЖЕНИЯ ДЛЯ МОДЕЛИ А

A.6 Остаточный член имеет нормальное распределение

Остаточный член *u* состоит из ряда факторов, явно не проявляющихся в уравнении регрессии, поэтому даже если мы не знаем о распределении этих факторов, мы можем предположить, что остаточный член имеет нормальное распределение.