

Лекция 5. Прогнозирование в регрессионных моделях ПР и МЛР

Введение: от объяснения к предсказанию

Мы научились оценивать параметры моделей, проверять гипотезы и интерпретировать коэффициенты. Теперь наша цель - использовать построенные модели для ответа на вопрос: «**Что будет?**». Прогнозирование - это одно из ключевых применений эконометрики в бизнесе, госуправлении и науке.

1. Виды прогнозов и этапы прогнозирования

1.1. Точечный и интервальный прогноз

- **Точечный прогноз** - это единственное число, наше «лучшее» предположение о будущем значении.
- **Интервальный прогноз** - это диапазон, в котором с заданной вероятностью будет находиться будущее значение. Он гораздо информативнее, так как учитывает неопределенность.

1.2. Этапы прогнозирования:

1. **Построение модели:** Оценивание параметров на основе исторических данных (выборки).
2. **Получение прогноза:** Подстановка значений регрессоров в оцененное уравнение.
3. **Оценка точности:** Расчет ошибки прогноза и построение прогнозного интервала.

2. Прогнозирование в парной линейной регрессии

Модель:

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

Оцененная модель:

$$\hat{Y}_i = b_0 + b_1 X_i$$

2.1. Точечный прогноз

Точечный прогноз для нового значения $X = X_0$ вычисляется прямой подстановкой: $\hat{Y}_0 = b_0 + b_1 X_0$

Пример: Если модель зарплаты $\hat{Y} = -20 + 5X$, то точечный прогноз зарплаты для человека с 17 годами образования: $\hat{Y}_0 = -20 + 5 \cdot 17 = 65$ (тыс. руб.).

2.2. Интервальный прогноз

Ошибка прогноза возникает из-за:

1. **Ошибки предсказания:** Случайная компонента ε_0 для нового наблюдения.
2. **Ошибки оценки:** Неточность в оценках b_0 и b_1 .

Учитывая это, **стандартная ошибка прогноза (SEF)** для нового наблюдения рассчитывается как:

$$SEF(\hat{Y}_0) = s * \sqrt{1 + 1/n + (X_0 - \bar{X})^2 / \sum (X_i - \bar{X})^2}$$

где s - стандартная ошибка регрессии ($s = \sqrt{RSS/(n - 2)}$).

Прогнозный интервал для индивидуального значения Y_0 с доверительной вероятностью $(1-\alpha)$:

$$\hat{Y}_0 \pm t_{(\alpha/2, n-2)} * SEF(\hat{Y}_0)$$

где $t_{(\alpha/2, n-2)}$ - критическое значение t-распределения.

3. Прогнозирование в множественной линейной регрессии

Модель:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i$$

Оцененная модель:

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \dots + b_k X_{ki}$$

3.1. Точечный прогноз

Точечный прогноз для нового вектора значений $X_0 = (1, X_{01}, X_{02}, \dots, X_{0k})$:

$$\hat{Y}_0 = b_0 + b_1 X_{01} + b_2 X_{02} + \dots + b_k X_{0k}$$

Пример: Модель зарплаты: $\hat{Y} = -20 + 4.5 * \text{Образование} + 1.2 * \text{Опыт} + 8.0 * \text{Пол}$

Прогноз для мужчины с 16 годами образования и 5-летним опытом:

$$\hat{Y}_0 = -20 + 4.5 * 16 + 1.2 * 5 + 8.0 * 1 = 66 \text{ (тыс. руб.)}$$

3.2. Интервальный прогноз

В матричной форме стандартная ошибка прогноза имеет вид:

$$SEF(\hat{Y}_0) = s * \sqrt{1 + X_0^T (X^T X)^{-1} X_0}$$

Прогнозный интервал для индивидуального значения Y_0 :

$$\hat{Y}_0 \pm t_{(\alpha/2, n-k-1)} * SEF(\hat{Y}_0)$$

где $n-k-1$ — число степеней свободы.

4. Особенности и проблемы прогнозирования

4.1. Экстраполяция

Экстраполяция - это прогнозирование за пределами диапазона данных, использовавшихся для построения модели.

Экстраполяция опасна, так как характер зависимости за пределами исследованного диапазона может измениться.

4.2. Стабильность параметров

Параметры модели, оцененные на исторических данных, могут оказаться нестабильными в будущем, если изменится экономический режим (например, финансовый кризис, изменение законодательства).

4.3. Мультиколлинеарность

Сильная мультиколлинеарность не влияет на точечный прогноз, но значительно увеличивает дисперсию прогноза, делая интервальные прогнозы чрезмерно широкими.

5. Оценка качества прогноза

5.1. Разделение выборки

Надежный метод оценки прогнозных свойств модели:

- **Обучающая выборка** (70-80% данных): для построения модели
- **Тестовая выборка** (20-30% данных): для оценки точности прогноза

5.2. Меры точности прогноза:

- **Средняя абсолютная ошибка (MAE):**

$$MAE = (1/m) * \sum |Y_i - \hat{Y}_i|$$

- **Средняя квадратическая ошибка (MSE):**

$$MSE = (1/m) * \sum (Y_i - \hat{Y}_i)^2$$

- **Корень из MSE (RMSE):**

$$RMSE = \sqrt{MSE} \text{ (интерпретируется в единицах } Y)$$

где m — количество наблюдений в тестовой выборке.

6. Практический пример: Прогноз продаж

Задача: спрогнозировать объем продаж на основе расходов на рекламу.

Оцененная модель (по данным за 3 года): Продажи = 100 + 15*Реклама

$$R^2 = 0.8, s = 20, n = 36, \bar{X} = 50, \sum (X_i - \bar{X})^2 = 1000$$

Точечный прогноз на следующий месяц при бюджете рекламы 70:

$$\hat{Y}_0 = 100 + 15 * 70 = 1150 \text{ (единиц)}$$

Интервальный прогноз (95% доверительная вероятность, $t_{0.0005, 33} \approx 2.03$):

- $SEF(\hat{Y}_0) = 20 * \sqrt{1 + 1/36 + (70 - 50)^2/1000} = 20 * \sqrt{1 + 0.028 + 0.4} \approx 20 * 1.2 = 24$

- **Прогнозный интервал:** $1150 \pm 2.03 * 24 = 1150 \pm 48.7$

- **Результат:** [1101.3; 1198.7]

Вывод: С вероятностью 95% объем продаж составит от 1101 до 1199 единиц.

Резюме

1. **Точечный прогноз** получается прямой подстановкой значений в оцененное уравнение регрессии.
2. **Интервальный прогноз** учитывает неопределенность, связанную со случайной ошибкой и оценкой параметров.
3. **Точность прогноза** снижается при экстраполяции и нестабильности параметров модели.
4. **Качество прогноза** следует проверять на тестовой выборке, не участвовавшей в построении модели.
5. Прогнозный интервал **шире** доверительного интервала для условного среднего, так как включает дополнительную неопределенность индивидуального наблюдения.

На следующей лекции: Мы перейдем к новой большой теме - **нарушениям предпосылок классической линейной модели** и методам их выявления и исправления.

Вопросы для самопроверки:

1. Почему прогнозный интервал всегда шире, чем доверительный интервал для условного математического ожидания?
2. Как изменится ширина прогнозного интервала, если увеличить объем выборки n ?

3. В чем главная опасность экстраполяции и можно ли ее полностью избежать на практике?