

Оглавление

Предисловие от издательства	16
Об авторах	17
Предисловие	18
Глава 1. Приступая к работе	21
1.1 Что можно прогнозировать?	22
1.2 Прогнозирование, цели и планирование	23
Прогнозирование	24
Цели	24
Планирование	24
Краткосрочные прогнозы	24
Среднесрочные прогнозы	24
Долгосрочные прогнозы	24
1.3 Определение того, что следует прогнозировать	24
1.4 Данные и методы прогнозирования	25
Прогнозирование временных рядов	26
Индикаторные переменные и прогнозирование временных рядов	27
1.5 Несколько тематических исследований	28
Случай 1	28
Случай 2	29
Случай 3	29
Случай 4	30
1.6 Базовые шаги в задаче прогнозирования	30
Шаг 1: определение задачи	30
Шаг 2: сбор информации	30
Шаг 3: предварительный (разведывательный) анализ	30
Шаг 4: выбор и подгонка моделей	31
Шаг 5: использование и оценивание модели прогнозирования	31
1.7 Перспектива статистического прогнозирования	31
1.8 Упражнения	33
1.9 Дальнейшее чтение	34
Библиография	34
Глава 2. Графика временных рядов	35
2.1 Объекты twibble	35
Переменная-индекс	35
Переменные-ключи	37
Работа с объектами twibble	38
Прочитать csv-файл и конвертировать его в twibble	41
Сезонный период	42

2.2	Графики времени.....	43
2.3	Закономерности временных рядов.....	45
	Тренд.....	45
	Сезонный.....	45
	Циклический.....	46
2.4	Сезонные графики.....	47
	Несколько сезонных периодов.....	48
2.5	Графики сезонных подрядов.....	50
	Пример: туристический отдых в Австралии.....	50
2.6	Диаграммы рассеяния.....	54
	Корреляция.....	56
	Матрицы диаграмм рассеяния.....	57
	Библиография.....	60
2.7	Графики лагов.....	60
2.8	Автокорреляция.....	61
	Тренды и сезонность на ACF-графиках.....	63
2.9	Белый шум.....	63
2.10	Упражнения.....	65
2.11	Дальнейшее чтение.....	67
	Библиография.....	68

Глава 3. Разложение временных рядов..... 69

3.1	Преобразования и корректировки.....	69
	Календарные корректировки.....	69
	Корректировки по численности населения.....	70
	Корректировки на инфляцию.....	71
	Математические преобразования.....	72
	Библиография.....	74
3.2	Компоненты временных рядов.....	75
	Пример: занятость в секторе розничной торговли США.....	75
	Сезонно скорректированные данные.....	79
3.3	Скольльзящие средние.....	80
	Сглаживание скользящей средней.....	80
	Скольльзящие средние скользящих средних.....	83
	Оценивание тренда-цикла с сезонными данными.....	85
	Пример: занятость в секторе розничной торговли США.....	85
	Взвешенные скользящие средние.....	86
3.4	Классическое разложение.....	87
	Аддитивное разложение.....	87
	Мультипликативное разложение.....	88
	Комментарии к классическому разложению.....	89
3.5	Методы, используемые официальными статистическими агентствами.....	89
	Метод X-11.....	90
	Метод SEATS.....	93
	Библиография.....	94
3.6	STL-разложение.....	95
	Библиография.....	97

3.7 Упражнения	97
3.8 Дальнейшее чтение	100
Библиография	100
Глава 4. Признаки временных рядов	101
4.1 Несколько простых статистик.....	101
4.2 Признаки ACF	102
4.3 Признаки STL	104
4.4 Другие признаки.....	107
4.5 Разведывательный анализ данных о туризме в Австралии.....	108
Библиография	114
4.6 Упражнения	114
4.7 Дальнейшее чтение	114
Библиография	114
Глава 5. Инструментарий прогнозиста	115
5.1 Чистый рабочий поток прогнозирования.....	115
Подготовка данных (очистка)	115
Построение графика данных (визуализация)	116
Определение модели (формулирование)	116
Тренировка модели (расчет).....	117
Проверка результативности модели (оценивание)	118
Генерирование прогнозов (прогнозирование)	118
5.2 Несколько простых методов прогнозирования.....	119
Метод среднего	120
Наивный метод.....	120
Сезонный наивный метод	121
Дрейфовый метод.....	122
Пример: ежеквартальное производство пива в Австралии.....	123
Пример: ежедневная цена акций Google на момент закрытия торгов...	124
5.3 Подогнанные значения и остатки	125
Подогнанные значения.....	125
Остатки.....	126
5.4 Диагностика остатков.....	127
Пример: прогнозирование цен акций Google на момент закрытия торгов	128
Тесты портманто на автокорреляцию	131
5.5 Распределительные прогнозы и предсказательные интервалы	133
Прогнозные распределения.....	133
Предсказательные интервалы	134
Одношаговые предсказательные интервалы	135
Многошаговые предсказательные интервалы	135
Эталонные методы	135
Предсказательные интервалы из бутстрапированных остатков	137
5.6 Прогнозирование с использованием преобразований.....	142
Предсказательные интервалы с преобразованиями	142
Поправки на смещение	142

5.7 Прогнозирование с разложением.....	144
Пример: занятость в секторе розничной торговли США.....	145
5.8 Оценивание точности точечного прогноза	147
Тренировочный и тестовый наборы	147
Функции для взятия подмножеств из временных рядов	148
Ошибки прогнозов	149
Ошибки, зависимые от шкалы	149
Процентные ошибки	149
Шкалированные ошибки.....	150
Примеры	151
Библиография	154
5.9 Оценивание точности распределительного прогноза	154
Квантильные баллы.....	154
Балл Винклера.....	155
Непрерывный рангово-вероятностный балл	156
Бесшкальные сравнения с использованием баллов навыков	157
Библиография	158
5.10 Перекрестная валидация временных рядов	158
Пример: точность прогнозного горизонта с помощью перекрестной валидации	161
5.11 Упражнения	162
5.12 Дальнейшее чтение	165
Библиография	165

Глава 6. Прогнозы на основе суждений 166

Библиография	167
6.1 Остерегайтесь ограничений.....	167
Библиография	168
6.2 Ключевые принципы	168
Четко и кратко формулировать задачу прогнозирования	168
Внедрять систематический подход.....	168
Документировать и обосновывать	169
Систематически оценивать прогнозы	169
Сегрегировать прогнозистов и пользователей	169
Пример: схема фармацевтических льгот (PBS)	170
6.3 Дельфийский метод.....	172
Эксперты и анонимность	173
Постановка задачи прогнозирования в дельфийском методе.....	173
Обратная связь.....	174
Итерация	174
Окончательные прогнозы.....	174
Ограничения и вариации.....	174
Куратор	174
Библиография	175
6.4 Прогнозирование по аналогии	175
Пример: разработка учебной программы средней школы	175
Структурированная аналогия.....	176

Библиография	176
6.5 Сценарное прогнозирование	177
6.6 Прогнозирование новых продуктов	177
Композитный прогноз торгового отдела	177
Мнение руководства	178
Намерения клиентов	178
Библиография	179
6.7 Сужденческие корректировки	179
Использовать корректировки экономно	179
Применять структурированный подход	180
Пример: комитет по прогнозированию туризма (TFC)	180
Библиография	181
6.8 Дальнейшее чтение	182
Библиография	182

Глава 7. Модели регрессии временных рядов 184

7.1 Линейная модель	184
Простая линейная регрессия	184
Пример: потребительские расходы в США	185
Множественная линейная регрессия	188
Пример: потребительские расходы в США	188
Допущения	190
7.2 Оценивание методом наименьших квадратов	191
Пример: потребительские расходы в США	191
Подогнанные значения	192
Качество подгонки	194
Пример: потребительские расходы в США	195
Стандартная ошибка регрессии	195
7.3 Оценивание регрессионной модели	195
График ACF остатков	196
Гистограмма остатков	196
Графики остатков в сопоставлении с предсказателями	197
Графики остатков в сопоставлении с подогнанными значениями	198
Выбросы и влиятельные наблюдения	199
Мнимая регрессия	200
7.4 Несколько полезных предсказателей	202
Тренд	202
Фиктивные переменные	203
Сезонные фиктивные переменные	203
Пример: ежеквартальное производство пива в Австралии	204
Переменные вмешательства	207
Торговые дни	207
Распределенные лаги	208
Пасха	208
Ряд Фурье	208
7.5 Отбор предсказателей	209
Скорректированный R^2	210

Перекрестная валидация	211
Информационный критерий Акаике	211
Исправленный информационный критерий Акаике	212
Байесов информационный критерий Шварца	212
Какую меру мы должны использовать?	212
Пример: потребление в США	213
Регрессия наилучшего подмножества	214
Пошаговая регрессия	214
Остерегайтесь статистического вывода после отбора предсказателей	215
Библиография	215
7.6 Прогнозирование с помощью регрессии	215
Антефактумные прогнозы в сопоставлении с постфактумными прогнозами	215
Пример: ежеквартальное производство пива в Австралии	216
Прогнозирование на основе сценариев	217
Строительство предсказательной регрессионной модели	218
Предсказательные интервалы	219
7.7 Нелинейная регрессия	221
Прогнозирование с нелинейным трендом	222
Пример: времена побед в Бостонском марафоне	222
7.8 Корреляция, каузация и прогнозирование	225
Корреляция – это не каузация	225
Прогнозирование с использованием коррелированных предсказателей	226
Мультиколлинеарность и прогнозирование	226
7.9 Матричная формулировка	227
Оценивание методом наименьших квадратов	228
Подогнанные значения и перекрестная валидация	228
Прогнозы и предсказательные интервалы	229
7.10 Упражнения	229
7.11 Дальнейшее чтение	232
Библиография	233

Глава 8. Экспоненциальное сглаживание 234

Библиография	234
8.1 Простое экспоненциальное сглаживание	235
Средневзвешенная форма	237
Компонентная форма	238
Плоские прогнозы	238
Оптимизация	239
Пример: экспорт из Алжира	239
8.2 Методы с трендом	241
Метод линейного тренда Хольта	241
Пример: население Австралии	242
Методы демпфируемого тренда	244
Пример: численность населения Австралии (продолжение)	245

Пример: используемость интернета	246
Библиография	248
8.3 Методы с сезонностью.....	249
Аддитивный метод Хольта–Винтерса	249
Мультипликативный метод Хольта–Винтерса.....	250
Пример: внутренние поездки с ночевкой в Австралии.....	250
Демпфирующий метод Хольта–Винтерса	254
Пример: метод Хольта–Винтерса с ежедневными данными	255
Библиография	255
8.4 Таксономия методов экспоненциального сглаживания.....	256
Библиография	257
8.5. Модели пространства состояний инноваций для экспоненциального сглаживания.....	257
ETS(A,N,N): простое экспоненциальное сглаживание с аддитивными ошибками.....	258
ETS(M,N,N): простое экспоненциальное сглаживание с мультипликативными ошибками.....	259
ETS(A,A, N): линейный метод Хольта с аддитивными ошибками	260
ETS(M,A, N): линейный метод Хольта с мультипликативными ошибками.....	260
Другие модели ETS	261
8.6 Оценивание и отбор модели	261
Оценивание моделей ETS	261
Отбор модели.....	262
Пример: гостевые ночи праздничных туристов в Австралии	263
Библиография	265
8.7 Прогнозирование с использованием моделей ETS	265
Предсказательные интервалы	266
Библиография	267
8.8 Упражнения.....	268
8.9 Дальнейшее чтение	271
Библиография	271

Глава 9. Модели ARIMA 272

9.1 Стационарность и дифференцирование	272
Дифференцирование	274
Модель случайного блуждания.....	275
Дифференцирование второго порядка.....	276
Сезонное дифференцирование	276
Тесты на единичные корни.....	279
Библиография	281
9.2 Нотация обратного сдвига	281
9.3 Авторегрессионные модели	282
9.4 Модели скользящих средних	283
9.5 Несезонные модели ARIMA	285
Пример: экспорт из Египта.....	286

Понимание моделей ARIMA	288
Графики ACF и PACF	288
9.6 Оценивание и выбор порядка	291
Оценивание максимального правдоподобия	291
Информационные критерии	291
9.7 Моделирование ARIMA в <i>fable</i>	292
Как работает функция ARIMA()?	292
Процедура моделирования	294
Пример: экспорт из Центральноафриканской Республики	296
Понимание констант в R	300
Построение графика характеристических корней	300
Библиография	301
9.8 Прогнозирование	301
Точечные прогнозы	301
Предсказательные интервалы	303
Библиография	304
9.9 Сезонные модели ARIMA	304
ACF/PACF	304
Пример: ежемесячная занятость в сфере досуга и гостиничного бизнеса в США	305
Пример: продажа кортикостероидных препаратов в Австралии	309
Оценивание на тестовом наборе	313
9.10 ARIMA в сопоставлении с ETS	314
Сравнение ARIMA() и ETS() на несезонных данных	316
Сравнение ARIMA() и ETS() на сезонных данных	317
9.11 Упражнения	321
9.12 Дальнейшее чтение	325
Библиография	326

Глава 10. Модели на основе динамической регрессии 327

10.1 Оценивание	328
Библиография	329
10.2 Регрессия с ошибками ARIMA с использованием <i>fable</i>	329
Пример: личное потребление и доход в США	330
10.3 Прогнозирование	333
Пример: личное потребление и доход в США	333
Пример: прогнозирование спроса на электроэнергию	334
10.4 Стохастические и детерминированные тренды	339
Пример: пассажиры воздушного транспорта Австралии	339
10.5 Динамическая гармоническая регрессия	342
Пример: расходы австралийцев на питание вне дома	343
Библиография	344
10.6 Лагированные предсказатели	344
Пример: телевизионная реклама и страховые котировки	345
10.7 Упражнения	348
10.8 Дальнейшее чтение	351
Библиография	351

Глава 11. Прогнозирование иерархических и сгруппированных временных рядов	352
11.1 Иерархические и сгруппированные временные ряды.....	353
Иерархические временные ряды	353
Пример: иерархия туризма в Австралии	354
Сгруппированные временные ряды	358
Пример: тюремное население Австралии	359
Смешанная иерархическая и групповая структура	362
11.2 Одноуровневые подходы.....	364
Подход «снизу вверх»	364
Пример: генерирование восходящих прогнозов	364
Рабочий поток прогнозирования агрегационных структур.....	366
Подходы «сверху вниз»	366
Средние исторические пропорции	366
Пропорции исторических средних	367
Прогнозные пропорции	367
Подход «от середины наружу»	369
Библиография	369
11.3 Увязывание прогнозов	369
Матричная нотация.....	369
Отображающие матрицы.....	370
Увязывание	371
Метод оптимальной увязки MinT.....	372
Библиография	374
11.4 Прогнозирование внутреннего туризма в Австралии	374
11.5 Увязанные распределительные прогнозы	377
Библиография	378
11.6 Прогнозирование тюремного населения Австралии	378
Библиография	383
11.7 Упражнения	383
11.8 Дальнейшее чтение	384
Библиография	384
Глава 12. Продвинутое прогнозирование	386
12.1 Сложная сезонность.....	386
STL с несколькими сезонными периодами	387
Динамическая гармоническая регрессия с несколькими сезонными периодами.....	390
Пример: спрос на электроэнергию	392
Библиография	395
12.2 Модель Prophet.....	396
Пример: ежеквартальное производство цемента	396
Пример: получасовой спрос на электроэнергию	398
Библиография	400
12.3 Векторные авторегрессии	400
Пример: модель VAR для прогнозирования потребления в США.....	403
Библиография	405

12.4 Нейросетевые модели.....	405
Архитектура нейронной сети	405
Нейросетевая авторегрессия	407
Пример: пятна на Солнце	407
Предсказательные интервалы	408
12.5 Бутстрапирование и бутстрап-агрегирование	410
Бутстрапирование временного ряда.....	410
Бутстрап-агрегированные прогнозы	412
Библиография	414
12.6 Упражнения	415
12.7 Дальнейшее чтение	415
Библиография	415

Глава 13. Некоторые практические вопросы

прогнозирования	417
13.1 Ежедневные, ежедневные и субдневные данные	417
Еженедельные данные	417
Ежедневные и субдневные данные	419
13.2 Временные ряды счетных данных.....	420
Пример: продажа фармацевтических препаратов	421
Библиография	423
13.3 Обеспечение того, чтобы прогнозы оставались в пределах допустимых значений.....	424
Положительные прогнозы	424
Прогнозы, ограниченные интервалом	425
13.4 Комбинации прогнозов	427
Прогнозные распределения комбинаций	429
Библиография	431
13.5 Предсказательные интервалы для агрегатов.....	431
13.6 Ретрополяция	432
13.7 Очень длинные и очень короткие временные ряды	433
Прогнозирование очень коротких временных рядов.....	433
Прогнозирование очень длинных временных рядов	435
13.8 Прогнозирование на тренировочном и тестовом наборах.....	435
Многошаговые прогнозы на тренировочных данных	436
Одношаговые прогнозы на тестовых данных	437
13.9 Работа с выбросами и пропущенными значениями.....	438
Выбросы	438
Пропущенные значения	441
13.10 Дальнейшее чтение.....	443
Библиография	443

Приложение А. Использование языка R

Инсталлирование R и RStudio.....	444
Примеры на R в этой книге.....	444
Начало работы с R.....	445

Приложение В. Для преподавателей.....	446
Решения упражнений.....	446
Слайды.....	446
Приложение С. Рецензии и отзывы	447
Отзывы коллег-преподавателей.....	447
Отзывы практикующих специалистов и студентов.....	448
Библиография	450

Об авторах



Роб Джей Хайндман (<https://robjhyndman.com/>) – профессор статистики и заведующий кафедрой эконометрики и бизнес-статистики Университета Монаш, Австралия. В 2005–2018 годах был главным редактором журнала *International Journal of Forecasting* (Международного журнала прогнозирования). Он является автором 6 книг и более 200 научных работ, а также избранным членом Австралийской академии наук, Академии социальных наук Австралии и Международного института прогнозистов. На протяжении более 35 лет Роб ведет активную консалтинговую практику, помогая сотням компаний и организаций в решении задач прогнозирования. За свои исследования, преподавание, консультирование и руководство выпускниками он по достоинству оценен наградами.



Джордж Атанасопулос (<https://research.monash.edu/en/persons/george-athanasopoulos>) – профессор, директор по образованию и заместитель заведующего кафедрой эконометрии и бизнес-статистики Университета Монаш, Австралия. Он также является президентом Международного института прогнозистов. За свои исследования и преподавание Джордж имеет множество наград и отличий. Он входит в состав редакционных советов *Journal of Travel Research* и *International Journal of Forecasting*.

Предисловие

Добро пожаловать в наш онлайн-учебник по прогнозированию.

Цель этого учебника – дать всестороннее введение в методы прогнозирования и представить достаточно информации о каждом методе, чтобы читатели имели возможность разумно их использовать. Мы не пытаемся подробно обсуждать теоретические детали, лежащие в основе каждого метода, хотя ссылки в конце каждой главы будут содержать многие из этих деталей.

Книга написана для трех аудиторий: (1) людей, которые занимаются прогнозированием в бизнесе, хотя у них, возможно, не было никакого формального обучения в этой области; (2) студентов бакалавриата, изучающих бизнес; (3) студентов по программе MBA, изучающих факультатив по прогнозированию. Мы сами используем этот учебник для студентов магистратуры и студентов третьего курса бакалавриата в Университете Монаш, Австралия.

В большинстве разделов мы предполагаем, что читатели лишь слегка знакомы с вводной статистикой и с алгеброй средней школы. Есть пара разделов, которые также требуют знания матриц, но они помечены.

В конце каждой главы мы приводим список «дальнейшего чтения». В целом эти списки включают предлагаемые учебники, которые обеспечивают более углубленное или подробное рассмотрение предмета. Там, где нет подходящего учебника, мы предлагаем журнальные статьи, содержащие больше информации.

На протяжении всей книги мы используем язык программирования R и хотим, чтобы студенты научились делать прогнозы с помощью R. Язык R бесплатен и доступен практически в любой операционной системе. Это прекрасный инструмент для любого статистического анализа, а не только для прогнозирования. Инструкции по установке и использованию R см. в приложении А «Использование языка R».

Все примеры на R в книге основаны на допущении, что вы сначала скачали пакет `fpp3`:

```
library(fpp3)
#> — Attaching packages — fpp3 0.4.0 —
#> ✓ tibble      3.1.6    ✓ tsibble      1.1.1
#> ✓ dplyr       1.0.8    ✓ tsibbledata  0.4.0
#> ✓ tidyr       1.2.0    ✓ feasts       0.2.2
#> ✓ lubridate   1.8.0    ✓ fable        0.3.1
#> ✓ ggplot2     3.3.5
#> — Conflicts — fpp3_conflicts —
#> ✗ lubridate::date() masks base::date()
#> ✗ dplyr::filter() masks stats::filter()
#> ✗ tsibble::intersect() masks base::intersect()
```

```
#> X tsibble::interval() masks lubridate::interval()
#> X dplyr::lag() masks stats::lag()
#> X tsibble::setdiff() masks base::setdiff()
#> X tsibble::union() masks base::union()
```

Это позволит загружать соответствующие наборы данных и прикреплять несколько пакетов, как указано выше. К ним относятся несколько пакетов `tidyverse` и пакеты для обработки временных рядов и прогнозирования в рамках «чистой» (`tidy`) структуры.

В приведенной выше распечатке также показаны версии пакетов, которые мы использовали при составлении этого издания книги. Некоторые примеры в книге не будут работать с более ранними версиями пакетов.

Наконец, в распечатке перечислены некоторые конфликты, показывающие, какая функция будет предпочтительнее, когда функция с тем же именем находится в нескольких пакетах.

Эта книга несколько отличается от других учебников по прогнозированию:

- она бесплатна и доступна онлайн, что делает ее доступной для широкой аудитории;
- в ней используется язык R, являющийся бесплатным, имеющим открытый исходный код и представляющий чрезвычайно мощное программное обеспечение;
- онлайн-версия постоянно обновляется. Вам не нужно ждать следующего выпуска, чтобы устранить ошибки или обсудить новые методы. Мы будем обновлять книгу часто;
- есть десятки примеров реальных данных, взятых из нашей собственной консалтинговой практики. Мы работали с сотнями предприятий и организаций, помогая им в вопросах прогнозирования, и этот опыт внес непосредственный вклад во многие приведенные здесь примеры, а также определил нашу общую философию прогнозирования;
- мы уделяем графическим методам больше внимания, чем большинство прогнозистов. Мы используем графики для проведения разведывательного анализа данных, анализа валидности подогнанных моделей и представления результатов прогнозирования.

Изменения в третьем издании

Наиболее важным изменением в 3-м издании книги является то, что мы используем пакеты `tsibble` и `fable` вместо пакета `forecast`. Это позволяет нам тесно интегрироваться с коллекцией пакетов `tidyverse`. Как следствие мы заменили многие образцы, чтобы воспользоваться преимуществами новых возможностей.

Мы также добавили некоторые новые материалы по признакам временных рядов и реорганизовали содержание таким образом, чтобы в главах 2–4 обсуждался разведывательный анализ временных рядов, перед тем как представлять какие-либо методы прогнозирования. Это связано с тем, что сначала мы должны хорошо понимать наши временные ряды, их закономерности и признаки, прежде чем пытаться строить какие-либо модели и делать какие-либо прогнозы.

Внимательные читатели предыдущих версий книги сообщали нам о любых обнаруженных ими опечатках или ошибках. Они немедленно обновлялись в онлайн-режиме. Без сомнения, мы допустили несколько новых ошибок, и мы исправим их онлайн, как только они будут обнаружены. Пожалуйста, продолжайте сообщать нам о таких вещах (<https://bit.ly/fpptypo>).

Если у вас есть вопросы об использовании пакетов R, обсуждаемых в этой книге, или о прогнозировании в целом, то, пожалуйста, задайте их на веб-сайте сообщества RStudio (<https://bit.ly/fpp3help>).

Счастливого прогнозирования!

Роб Джей Хайндман и Джордж Атанасопулос

Май 2021 года

Для цитирования онлайн-версии этой книги, пожалуйста, используйте следующее:

Хайндман Р. Дж. и Атанасопулос Дж. (2021) Прогнозирование: принципы и практика. 3-е изд. OTexts: Мельбурн, Австралия. OTexts.com/fpp3.

Эта онлайн-версия книги в последний раз обновлялась 15 июня 2022 года.

Печатная версия книги (доступна на Amazon, <https://otexts.com/fpp3/buy-a-print-version.html>) в последний раз обновлялась 31 мая 2021 года.

Приступая к работе

Предсказание очаровывало людей на протяжении тысячелетий, иногда считаясь признаком божественного вдохновения, а иногда рассматриваясь как преступная деятельность. Еврейский пророк Исаия написал примерно в 700 году до нашей эры:

Скажите, что произойдёт в будущем, и мы будем знать, что вы боги.
(Исаия 41:23)

Сто лет спустя, в древнем Вавилоне, предсказатели предвещали будущее, основываясь на распределении личинок по гнилой овечьей печени. Примерно в то же время люди, желающие получить прогнозы, отправлялись в Дельфы в Греции, чтобы посоветоваться с Оракулом, который давал свои предсказания, находясь в состоянии алкогольного опьянения парами этилена. Предсказателям пришлось труднее при императоре Константине, который в 357 году нашей эры издал указ, запрещающий кому бы то ни было «справляться у провидца, математика или предсказателя... Пусть любопытство к предсказанию будущего умолкнет навсегда». Аналогичный запрет на предсказания был введен в Англии в 1736 году, когда мошенничество путем взимания денег за предсказания стало преступлением. Наказанием было трехмесячное тюремное заключение с каторжными работами!

Переменчивая судьба прогнозистов возникает из-за того, что хорошие прогнозы могут казаться почти волшебными, в то время как плохие прогнозы могут быть опасными. Рассмотрим следующие известные предсказания о вычислительной технике.

- Я думаю, что на мировом рынке существует, может быть, пять компьютеров. (Председатель правления IBM, 1943)
- Компьютеры в будущем могут весить не более 1.5 т. (Популярная механика, 1949)
- Нет никаких причин, по которым кто-то хотел бы иметь компьютер в своем доме. (Президент компании DEC, 1977 года)

Последний из них был сделан всего за три года до того, как IBM выпустила первый персональный компьютер. Неудивительно, что вы больше не можете купить компьютер DEC. Очевидно, что прогнозирование – это сложная деятельность, и компании, которые делают это хорошо, имеют большое преимущество перед теми, чьи прогнозы терпят неудачу.

В этой книге мы разведем наиболее надежные методы генерирования прогнозов. Акцент будет сделан на методах, которые можно воспроизводить и тестировать и которые, как было доказано, работают на практике.

1.1 Что можно прогнозировать?

Прогнозирование требуется во многих ситуациях: для принятия решения о строительстве еще одной электростанции в ближайшие пять лет требуются прогнозы будущего спроса; для составления расписания персонала в центре обработки звонков на следующей неделе требуются прогнозы объемов звонков; для пополнения запасов требуются прогнозы потребностей в запасах. Прогнозы могут требоваться на несколько лет вперед (в случае капитальных вложений) или всего на несколько минут (для маршрутизации телекоммуникаций). Какими бы ни были обстоятельства или временные горизонты, прогнозирование является важным подспорьем для эффективного и результативного планирования.

Некоторые вещи легче предсказывать, чем другие. Время восхода солнца завтра утром можно точно предсказать. С другой стороны, завтрашние лотерейные номера невозможно предсказать с какой-либо точностью. Предсказуемость события или количества зависит от нескольких факторов, включая:

- 1) насколько хорошо мы понимаем способствующие этому факторы;
- 2) каков объем доступных данных;
- 3) как будущее похоже на прошлое;
- 4) могут ли прогнозы повлиять на то, что мы пытаемся спрогнозировать.

Например, краткосрочные прогнозы спроса на электроэнергию в жилых домах могут быть очень точными, поскольку обычно удовлетворяются все четыре условия.

1. У нас есть хорошее представление о способствующих факторах: спрос на электроэнергию в основном зависит от температуры, с меньшим влиянием календарных изменений, таких как праздники, и экономических условий.
2. Обычно есть данные о спросе на электроэнергию за несколько лет и данные о погодных условиях за многие десятилетия.
3. В случае краткосрочного прогнозирования (до нескольких недель) можно с уверенностью предположить, что поведение спроса будет аналогичным тому, что наблюдалось в прошлом.
4. Для большинства бытовых потребителей цена на электроэнергию не зависит от спроса, и поэтому прогнозы спроса практически не влияют на поведение потребителей.

При условии что у нас есть навыки разработки хорошей модели, связывающей спрос на электроэнергию с ключевыми факторами, прогнозы могут быть удивительно точными.

С другой стороны, при прогнозировании курсов обмена валют удовлетворяется только одно из условий: есть достаточно данных. Однако у нас ограниченное понимание факторов, влияющих на обменные курсы, будущее вполне может отличаться от прошлого, если в одной из стран произойдет финансовый или политический кризис, а прогнозы обменного курса оказывают прямое влияние на сами курсы. Если есть широко разрекламированные прогнозы о том, что обменный курс будет расти, то люди немедленно скорректируют цену, которую они готовы заплатить, и поэтому прогнозы самореализуются.

В некотором смысле обменные курсы становятся их собственными прогнозами. Это пример «гипотезы эффективного рынка». Следовательно, прогнозирование повышения или падения завтрашнего обменного курса примерно так же предсказуемо, как прогнозирование выпадения подброшенной монеты орлом или решкой. В обеих ситуациях вы будете правы примерно в 50 % случаев, что бы вы ни прогнозировали. В подобных ситуациях прогнозисты должны осознавать свои собственные ограничения и не требовать больше, чем возможно.

Часто при прогнозировании ключевым шагом является знание того, когда что-то можно точно спрогнозировать, а когда прогнозы будут не лучше подбрасывания монетки. Хорошие прогнозы отражают подлинные закономерности и взаимосвязи, существующие в исторических данных, но не повторяют прошлые события, которые больше не повторятся. В этой книге мы узнаем, как отличать случайную флуктуацию в прошлых данных, которую следует игнорировать, от подлинной закономерности, которую следует смоделировать и экстраполировать.

Многие люди ошибочно полагают, что в меняющихся условиях прогнозы невозможны. Любая среда меняется, и хорошая модель прогнозирования улавливает то, как все меняется. Прогнозы редко строят на том, что окружающая среда остается неизменной. Обычно исходят из того, что характер изменения окружающей среды будет продолжаться и в будущем. То есть сильно изменчивая среда будет оставаться очень изменчивой; бизнес с колеблющимися объемами продаж будет продолжать иметь колеблющиеся объемы продаж, а экономика, которая пережила подъемы и спады, будет продолжать переживать подъемы и спады. Модель прогнозирования предназначена для того, чтобы улавливать то, как движутся события, а не только то, где они находятся. Как сказал Авраам Линкольн: «Если бы мы могли прежде узнать, где мы находимся и куда мы движемся, то мы могли бы лучше судить, что делать и как это делать».

Ситуации прогнозирования сильно различаются по своим временным горизонтам, факторам, определяющим фактически результаты, типам моделей данных и многим другим аспектам. Методы прогнозирования бывают простыми, такими как использование самого последнего наблюдения в качестве прогноза (который называется наивным методом), или очень сложными, такими как нейронные сети и эконометрические системы одновременных уравнений. Иногда данные вообще будут отсутствовать. Например, у нас может возникнуть потребность спрогнозировать продажи нового продукта в первый год его выпуска, но очевидно, что данных для работы нет. В подобных ситуациях мы используем сужденческое (или рассудочное) прогнозирование, рассмотренное в главе 6. Выбор метода зависит от имеющихся данных и предсказуемости прогнозируемой величины.

1.2 ПРОГНОЗИРОВАНИЕ, ЦЕЛИ И ПЛАНИРОВАНИЕ

Прогнозирование – это обычная статистическая задача в бизнесе, где оно помогает принимать обоснованные решения о планировании производства, транспортировки и персонала, а также служит руководством для долгосрочного стратегического планирования. Однако бизнес-прогнозирование часто выполняется слабо, и его нередко путают с планированием и целями. Это три разные вещи.

Прогнозирование

заключается в максимально точном предсказании будущего с учетом всей доступной информации, включая исторические данные и знания о любых будущих событиях, которые могут повлиять на прогнозы.

Цели

– это то, что вы хотели бы, чтобы произошло. Цели должны быть увязаны с прогнозами и планами, но это происходит не всегда. Слишком часто цели ставятся без какого-либо плана их достижения и без каких-либо прогнозов относительно того, насколько они реалистичны.

Планирование

– это реакция на прогнозы и цели. Планирование включает в себя определение надлежащих действий, необходимых для того, чтобы ваши прогнозы соответствовали вашим целям.

Прогнозирование должно быть неотъемлемой частью деятельности руководства по принятию решений, поскольку оно может играть важную роль во многих областях деятельности компании. Современным организациям требуются краткосрочные, среднесрочные и долгосрочные прогнозы, в зависимости от конкретного применения.

Краткосрочные прогнозы

необходимы для планирования персонала, производства и транспортировки. В рамках процесса планирования часто также требуются прогнозы спроса.

Среднесрочные прогнозы

необходимы для определения будущих потребностей в ресурсах, для закупки сырья, найма персонала или покупки машин и оборудования.

Долгосрочные прогнозы

используются в стратегическом планировании. Такие решения должны принимать во внимание рыночные возможности, факторы окружающей среды и внутренние ресурсы.

От организации требуется разработка системы прогнозирования, которая включает в себя несколько подходов к прогнозированию неопределенных событий. Такие системы прогнозирования требуют развития экспертных знаний в области выявления задач прогнозирования, применения ряда методов прогнозирования, выбора подходящих методов для каждой задачи, а также оценивания и совершенствования методов прогнозирования с течением времени. Также важно иметь сильную организационную поддержку по использованию формальных методов прогнозирования, если мы хотим, чтобы они применялись успешно.

1.3 ОПРЕДЕЛЕНИЕ ТОГО, ЧТО СЛЕДУЕТ ПРОГНОЗИРОВАТЬ

На ранних стадиях проекта прогнозирования необходимо принять решения о том, что следует прогнозировать. Например, если прогнозы требуются для

изделий в производственной среде, то необходимо спросить, нужны ли прогнозы:

- 1) для каждой продуктовой линейки или для групп продуктов?
- 2) для каждой торговой точки, или для торговых точек, сгруппированных по регионам, или только для суммарного объема продаж?
- 3) еженедельные данные, ежемесячные данные или годовые данные?

Также необходимо учитывать прогнозный горизонт. Потребуется ли прогнозы на один месяц вперед, на шесть месяцев или на десять лет? Потребуется различные типы моделей, в зависимости от того, какой прогнозный горизонт является наиболее важным.

Как часто требуются прогнозы? Прогнозы, которые необходимо производить часто, лучше делать с помощью автоматизированной системы, чем с помощью методов, требующих тщательной ручной работы.

Стоит потратить время на разговоры с людьми, которые будут использовать прогнозы, дабы убедиться, что вы понимаете их потребности и то, как следует использовать прогнозы, прежде чем приступать к обширной работе по генерированию прогнозов.

После того как было определено, какие прогнозы требуются, необходимо найти или собрать данные, на которых прогнозы будут основываться. Необходимые для прогнозирования данные, возможно, уже существуют. В наши дни регистрируется много данных, и задача прогнозиста нередко состоит в том, чтобы определить, где и как хранятся необходимые данные. Данные могут включать данные о продажах компании, исторический спрос на продукт или уровень безработицы в географическом регионе. Значительная часть времени прогнозиста может быть потрачена на поиск и сопоставление имеющихся данных до разработки подходящих методов прогнозирования.

1.4 ДАННЫЕ И МЕТОДЫ ПРОГНОЗИРОВАНИЯ

Подходящие методы прогнозирования во многом зависят от имеющихся данных.

Если данных нет или если располагаемые данные не имеют отношения к прогнозам, то необходимо использовать методы **качественного прогнозирования**. Эти методы не являются чисто догадками – существуют хорошо разработанные структурированные подходы к получению хороших прогнозов без использования исторических данных. Эти методы обсуждаются в главе 6.

Количественное прогнозирование может применяться при соблюдении двух условий:

- 1) доступна числовая информация о прошлом;
- 2) разумно допустить, что некоторые аспекты прошлых закономерностей продолжатся и в будущем.

Существует широкий спектр методов количественного прогнозирования, часто разработанных в рамках конкретных дисциплин для конкретных целей. Каждый метод имеет свои собственные свойства, точность и затраты, которые необходимо учитывать при выборе конкретного метода.

В большинстве задач количественного прогнозирования используются либо данные временных рядов (собираемые через регулярные промежутки време-

ни), либо данные поперечного сечения (собранные в один момент времени). В этой книге мы занимаемся прогнозированием будущих данных и концентрируемся на области временных рядов.

Прогнозирование временных рядов

Примеры данных временных рядов таковы:

- годовая прибыль Google;
- квартальные результаты продаж Amazon;
- месячное количество осадков;
- еженедельные розничные продажи;
- ежедневные цены на акции IBM;
- почасовая потребность в электроэнергии;
- объемы 5-минутного дорожного движения на автостраде;
- данные с метками времени о биржевых транзакциях.

Все, что наблюдается последовательно с течением времени, является временным рядом. В этой книге мы будем рассматривать только те временные ряды, которые наблюдаются через регулярные промежутки времени (например, еже часно, ежедневно, еженедельно, ежемесячно, ежеквартально, ежегодно). Временные ряды с неравномерным интервалом также могут встречаться, но они выходят за рамки этой книги.

При прогнозировании данных временных рядов цель состоит в том, чтобы оценить, как последовательность наблюдений будет продолжаться в будущем. На рис. 1.1 показано ежеквартальное производство пива в Австралии с 2000-го по второй квартал 2010 года.

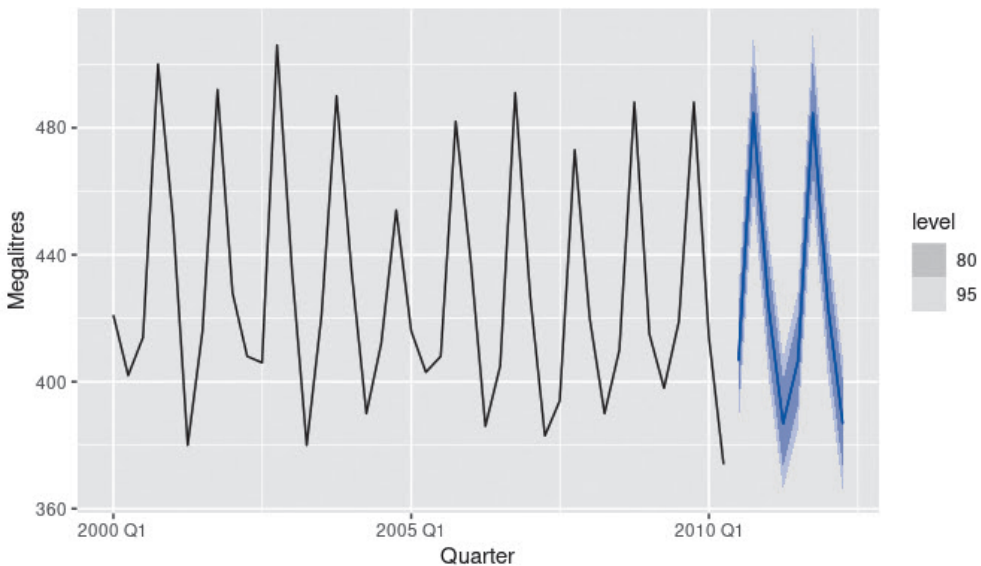


Рисунок 1.1: Квартальное производство пива в Австралии: 2000Q1–2010Q2, с прогнозами на два года

Синие линии показывают прогнозы на ближайшие два года. Обратите внимание, как прогнозы отражают сезонную закономерность, наблюдаемую в исторических данных, и воспроизводят ее в течение следующих двух лет. Темная заштрихованная область показывает 80%-ные предсказательные интервалы. То есть ожидается, что каждое будущее значение будет находиться в темном заштрихованном участке с вероятностью 80 %. Участок со светлым оттенком показывает 95%-ные предсказательные интервалы. Эти предсказательные интервалы являются полезным способом изображения неопределенности в прогнозах. В данном случае ожидается, что прогнозы будут точными, и, следовательно, предсказательные интервалы довольно узки.

В простейших методах прогнозирования временных рядов используется информация только о переменной, подлежащей прогнозированию, не пытаясь обнаруживать факторы, влияющие на ее поведение. Поэтому они будут экстраполировать трендовые и сезонные закономерности, но игнорировать всю остальную информацию, такую как маркетинговые инициативы, активность конкурентов, изменения экономических условий и т. д.

Методы разложения полезны для изучения трендовых и сезонных закономерностей во временных рядах; они обсуждаются в главе 3. Используемые для прогнозирования популярные модели временных рядов включают модели экспоненциального сглаживания и модели ARIMA, обсуждаемые соответственно в главах 8 и 9.

Индикаторные переменные и прогнозирование временных рядов

При прогнозировании временных рядов широко используются предсказательные переменные. Например, предположим, что мы хотим спрогнозировать почасовую потребность в электроэнергии (ED) в жарком регионе в летний период. Модель с предсказательными переменными может иметь следующую форму:

$$ED = f(\text{текущая температура, прочность экономики, население, время дня, день недели, ошибка}).$$

Взаимосвязь не является точной – в спросе на электроэнергию всегда будут изменения, учесть которые предсказательные переменные не смогут. Член «ошибка» справа учитывает случайную вариацию и влияние релевантных переменных, которые не включены в модель. Мы называем эту модель **объяснительной моделью**, потому что она помогает объяснять факторы, которые вызывают изменение спроса на электроэнергию.

Поскольку данные спроса на электроэнергию формируют временные ряды, мы также могли бы использовать модель временных рядов для прогнозирования. В этом случае подходящее уравнение прогнозирования временных рядов имеет форму

$$ED_{t+1} = f(ED_t, ED_{t-1}, ED_{t-2}, ED_{t-3}, \dots, \text{ошибка}),$$

где t – это текущий, $t + 1$ – следующий час, $t - 1$ – предыдущий час, $t - 1$ – два часа назад и т. д. Здесь прогнозирование будущего основано на прошлых значениях переменной, но не на внешних переменных, которые могут влиять

на систему. Опять же, член «ошибка» справа допускает случайную вариацию и влияние релевантных переменных, которые не включены в модель.

Кроме того, существует и третий тип моделей, который сочетает в себе особенности двух вышеупомянутых моделей. Например, ее можно задать следующей ниже формулой:

$$ED_{t+1} = f(ED_t, \text{текущая температура, время дня, день недели, ошибка}).$$

Указанные выше типы «смешанных моделей» в разных дисциплинах получили различные названия. Они называются моделями динамической регрессии, моделями панельных данных, продольными моделями, моделями передаточных функций и моделями линейных систем (при условии что f является линейной). Указанные модели обсуждаются в главе 10.

Объяснительная модель полезна тем, что она включает информацию о других переменных, а не только исторические значения переменной, подлежащей прогнозированию. Однако существует несколько причин, по которым прогнозист выберет именно модель временных рядов, а не объяснительную или смешанную модель. Во-первых, система может быть непонятна, и даже если бы она была понятной, бывает чрезвычайно трудно измерить взаимосвязи, которые, как предполагается, управляют ее поведением. Во-вторых, для того чтобы иметь возможность прогнозировать интересующую переменную, необходимо знать или прогнозировать будущие значения различных предсказателей, а это бывает слишком сложно делать. В-третьих, главная озабоченность может быть сконцентрирована только на том, чтобы предсказывать, что произойдет, а не на том, чтобы узнавать, почему это происходит. Наконец, модель временных рядов может давать более точные прогнозы, чем объяснительная или смешанная модель.

Модель, которая будет использоваться при прогнозировании, зависит от имеющихся ресурсов и данных, точности конкурирующих моделей и способа использования модели прогнозирования.

1.5 НЕСКОЛЬКО ТЕМАТИЧЕСКИХ ИССЛЕДОВАНИЙ

Следующие четыре примера взяты из нашей консалтинговой практики и демонстрируют различные типы ситуаций прогнозирования и связанные с ними проблемы, которые нередко возникают.

Случай 1

Клиентом была крупная компания, производящая одноразовую посуду, такую как салфетки и бумажные тарелки. Им нужны были прогнозы по каждому из сотен пунктов каждый месяц. Данные временных рядов показали целый ряд закономерностей: некоторые с трендами, некоторые сезонные, а некоторые ни с тем, ни с другим. В то время они использовали свое собственное программное обеспечение, написанное собственными силами, но оно часто давало прогнозы, которые не выглядели разумными. Использовались следующие методы:

- среднее значение данных за последние 12 месяцев;
- среднее значение данных за последние 6 месяцев;
- прогноз на основе прямой регрессии за последние 12 месяцев;

- прогноз на основе прямой регрессии за последние 6 месяцев;
- прогноз, полученный с помощью прямой линии через последнее наблюдение с наклоном, равным среднему наклону линий, соединяющих значения за прошлый и текущий годы;
- прогноз, полученный с помощью прямой линии через последнее наблюдение с наклоном, равным среднему наклону линий, соединяющих значения за прошлый и текущий годы, где среднее значение берется только за последние 6 месяцев.

Им нужно было, чтобы мы рассказали им, что происходит не так, и модифицировали программное обеспечение, чтобы оно давало более точные прогнозы. Программное обеспечение было написано на языке COBOL, что затрудняло выполнение каких-либо сложных численных вычислений.

Случай 2

В данном случае заказчиком выступило федеральное правительство Австралии, которому необходимо было спрогнозировать годовой бюджет Программы фармацевтических льгот (PBS). PBS предоставляет субсидии на многие фармацевтические продукты, продаваемые в Австралии, и расходы зависят от того, что люди покупают в течение года. В 2009 году суммарные расходы составили около 7 млрд австралийских долларов и занижались почти на 1 млрд долларов за каждые два года, прежде чем нас попросили помочь в разработке более точного подхода к прогнозированию.

Для того чтобы спрогнозировать суммарные расходы, необходимо спрогнозировать объемы продаж сотен групп фармацевтических продуктов, используя ежемесячные данные. Почти все группы имеют трендовые и сезонные закономерности. Объемы продаж многих групп резко увеличиваются или уменьшаются из-за изменений в том, какие лекарства субсидируются. Расходы для многих групп также претерпевают внезапные изменения из-за появления более дешевых препаратов конкурентов.

Таким образом, нам нужно было найти метод прогнозирования, который учитывал бы тренды и сезонность, если они присутствовали, и в то же время был устойчив к внезапным изменениям в базовых моделях. Кроме того, необходимо было иметь возможность автоматического применения к большому числу временных рядов.

Случай 3

Крупная компания, занимающаяся автопарком, обратилась к нам с просьбой помочь им спрогнозировать стоимость перепродажи автомобилей. Они покупают новые автомобили, сдают их в аренду на три года, а затем продают. Более точные прогнозы стоимости продаж транспортных средств означали бы лучший контроль над прибылью; понимание того, что влияет на стоимость перепродажи, может позволить разработать политику лизинга и продаж с целью максимизации прибыли.

В то время стоимость перепродажи прогнозировалась группой специалистов. К сожалению, они рассматривали любую статистическую модель как угрозу своей работе и отказывались сотрудничать в предоставлении информации. Тем не менее компания предоставила большое количество данных о предыдущих автомобилях и их возможной стоимости при перепродаже.

Случай 4

В этом проекте нам нужно было разработать модель для прогнозирования еженедельного пассажиропотока на основных внутренних маршрутах для одной из ведущих авиакомпаний Австралии. Компании требовались прогнозы числа пассажиров для каждого основного внутреннего маршрута и для каждого класса пассажиров (эконом-класс, бизнес-класс и первый класс). Компания предоставила еженедельные данные о трафике за предыдущие шесть лет.

На число авиапассажиров влияют школьные каникулы, крупные спортивные мероприятия, рекламные кампании, поведение конкурентов и т. д. Школьные каникулы в разных городах Австралии часто не совпадают, а спортивные мероприятия иногда переносятся из одного города в другой. В период исторических данных произошла крупная забастовка пилотов, во время которой в течение нескольких месяцев не было движения. Также была запущена и закрыта новая авиакомпания со сниженными ценами. Ближе к концу исторических данных авиакомпания опробовала перераспределение некоторых мест эконом-класса в бизнес-класс, а некоторых мест бизнес-класса – в первый класс. Однако через несколько месяцев классификация мест вернулась к первоначальному распределению.

1.6 БАЗОВЫЕ ШАГИ В ЗАДАЧЕ ПРОГНОЗИРОВАНИЯ

Задача прогнозирования обычно включает в себя пять базовых шагов.

Шаг 1: определение задачи

Нередко это самая сложная часть прогнозирования. Тщательное определение задачи требует понимания того, как прогнозы будут использоваться, кому прогнозы требуются и как функция прогнозирования вписывается в организацию, требующую прогнозов. Прогнозисту необходимо потратить время на общение со всеми, кто будет заниматься сбором данных, ведением баз данных и использованием прогнозов для будущего планирования.

Шаг 2: сбор информации

Всегда требуется по крайней мере два вида информации: (а) статистические данные и (б) накопленный опыт людей, которые собирают данные и используют прогнозы. Нередко бывает трудно получить достаточное количество исторических данных, чтобы иметь возможность выполнить подгонку хорошей статистической модели. В этом случае можно использовать методы сужденческого прогнозирования, описанные в главе 6. Иногда старые данные будут менее полезны из-за структурных изменений в прогнозируемой системе; тогда мы можем выбрать использование только самых последних данных. Однако помните, что хорошие статистические модели будут учитывать эволюционные изменения в системе; не выбрасывайте хорошие данные без необходимости.

Шаг 3: предварительный (разведывательный) анализ

Всегда следует начинать с построения графиков данных. Существуют ли неуклонные закономерности? Есть ли какой-то существенный тренд? Важна ли

сезонность? Есть ли подтверждающие данные о наличии бизнес-циклов? Есть ли какие-либо отклонения в данных, которые должны быть объяснены теми, кто обладает экспертными знаниями? Насколько сильны взаимосвязи между переменными, доступными для анализа? Для проведения этого анализа были разработаны различные инструменты. Они обсуждаются в главах 2 и 3.

Шаг 4: выбор и подгонка моделей

Наилучшая для использования модель зависит от наличия исторических данных, силы взаимосвязей между прогнозной переменной и любыми объясняющими переменными, а также от способа использования прогнозов. Обычно сравнивают две или три потенциальные модели. Каждая модель сама по себе является искусственной конструкцией, основанной на наборе допущений (явных и неявных), и обычно включает в себя один или несколько параметров, которые должны быть оценены с использованием известных исторических данных. Мы обсудим регрессионные модели (глава 7), методы экспоненциального сглаживания (глава 8), модели ARIMA Бокса–Дженкинса (глава 9), динамические регрессионные модели (глава 10), иерархическое прогнозирование (глава 11) и несколько продвинутых методов, включая нейронные сети и векторную авторегрессию (глава 12).

Шаг 5: использование и оценивание модели прогнозирования

После выбора модели и оценивания ее параметров модель используется для продуцирования прогнозов. Результативность модели можно надлежаще оценить только после того, как станут доступны данные за прогнозируемый период. Был разработан ряд методов, помогающих оценивать точность прогнозов. Кроме того, при использовании прогнозов и принятии мер в соответствии с ними есть организационные проблемы. Краткое обсуждение некоторых из этих вопросов приведено в главе 5. При использовании модели прогнозирования на практике возникает целый ряд практических проблем, таких как то, как обрабатывать пропущенные значения и выбросы или как работать с короткими временными рядами. Они обсуждаются в главе 13.

1.7 ПЕРСПЕКТИВА СТАТИСТИЧЕСКОГО ПРОГНОЗИРОВАНИЯ

То, что мы пытаемся спрогнозировать, неизвестно (иначе мы бы не прогнозировали это), и поэтому о нем можно думать как о случайной величине. Например, суммарный объем продаж за следующий месяц может принимать различные возможные значения, и пока мы не сложим фактические продажи в конце месяца, мы не узнаем, каким будет значение. Поэтому, пока мы не узнаем объем продаж за следующий месяц, это случайная величина.

Поскольку следующий месяц относительно близок, у нас обычно есть хорошее представление о том, какими могут быть вероятные показатели продаж. С другой стороны, если мы прогнозируем продажи за тот же месяц в следующем году, то возможные значения, которые он может принять, гораздо более изменчивы. В большинстве ситуаций прогнозирования вариаци-

ции, связанные с тем, что мы прогнозируем, будут уменьшаться по мере приближения события. Другими словами, чем дальше мы прогнозируем, тем больше у нас сомнений.

Можно представить себе множество вариантов возможного будущего, каждое из которых дает разную стоимость того, что мы хотим спрогнозировать. На рис. 1.2 черным цветом показано общее число международных прибытий в Австралию с 1980 по 2015 год. Также показаны десять вариантов возможного будущего на 2016–2025 годы.

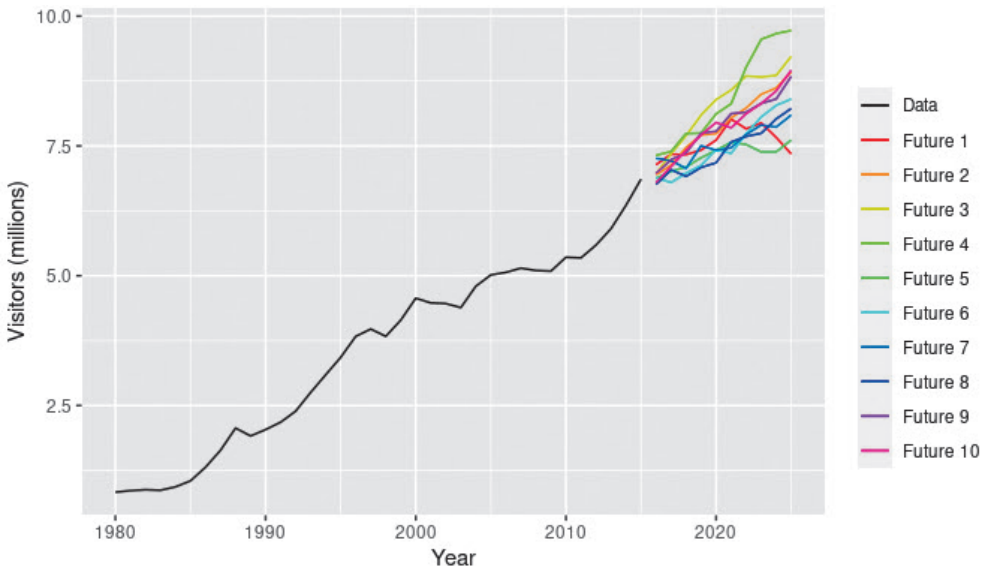


Рисунок 1.2: Общее число иностранных туристов, посетивших Австралию (1980–2015 гг.), а также десять вариантов возможного будущего

Когда мы получаем прогноз, мы оцениваем середину диапазона возможных значений, которые случайная величина может принимать. Нередко прогноз сопровождается предсказательным интервалом, дающим диапазон значений, которые случайная величина может принимать с относительно высокой вероятностью. Например, 95%-ный предсказательный интервал содержит диапазон значений, который должен включать фактическое будущее значение с вероятностью 95 %.

Вместо нанесения на график отдельных вариантов возможного будущего, как показано на рис. 1.2, мы обычно показываем эти предсказательные интервалы. На рис. 1.3 показаны интервалы между 80 % и 95 % в отношении будущих иностранных посетителей Австралии. Синяя линия – это среднее значение возможных будущих значений, которые мы называем **точечными прогнозами**.

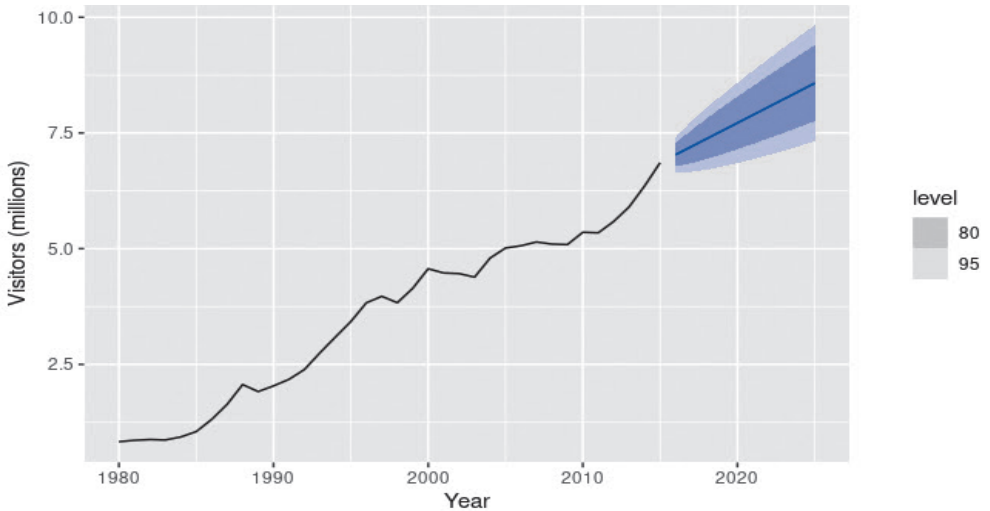


Рисунок 1.3: Общее число иностранных посетителей Австралии (1980–2015 гг.) вместе с прогнозами на 10 лет и 80%-ным и 95%-ным предсказательными интервалами

Мы будем использовать индекс t для времени. Например, y_t будет обозначать наблюдение в момент времени t . Предположим, мы обозначим всю информацию, которую мы наблюдали, как I , и хотим спрогнозировать y_t . Тогда мы записываем $y_t|I$, что означает «случайную величину y_t с учетом того, что мы знаем в I ». Набор значений, которые может принимать эта случайная величина, наряду с их относительными вероятностями называется «распределением вероятностей» случайной величины $y_t|I$. В прогнозировании мы называем его **прогноznым распределением**.

Когда мы говорим о «прогнозе», мы обычно имеем в виду среднее значение прогнозного распределения, и мы ставим «шляпку» над y , чтобы это показать. Таким образом, мы записываем прогноз y_t как $\hat{y}_{t|t-1}$, что означает среднее из возможных значений, которые y_t может принимать, учитывая все, что мы знаем.

Нередко бывает полезно точно указывать, какую информацию мы использовали при расчете прогноза. Тогда мы запишем, например, $\hat{y}_{t|t-1}$, чтобы обозначить прогноз y_t , принимая во внимание все предыдущие наблюдения (y_1, \dots, y_{t-1}). Аналогичным образом $\hat{y}_{T+h|T}$ означает прогноз y_{T+h} с учетом y_1, \dots, y_T (т. е. прогноз на h шагов вперед с учетом всех наблюдений вплоть до времени T).

1.8 УПРАЖНЕНИЯ

1. Для случаев 3 и 4 в разделе 1.5 перечислите возможные предсказательные переменные, которые могут быть полезны, при условии наличия соответствующих данных.
2. Для случая 3 в разделе 1.5 опишите пять шагов прогнозирования в контексте этого проекта.

1.9 ДАЛЬНЕЙШЕЕ ЧТЕНИЕ

Армстронг (2001, <https://otexts.com/fpp3/intro-reading.html#ref-Armstrong01>) охватывает всю область прогнозирования, причем каждая глава написана разными экспертами. Временами он очень самоуверен (и мы не со всем с ним согласны), но он полон превосходных общих советов по решению задач прогнозирования.

Орд и соавт. (2017, <https://otexts.com/fpp3/intro-reading.html#ref-Ord2017>) – это учебник по прогнозированию, охватывающий несколько тем, что и в этой книге, но с другим акцентом и не ориентированный на какую-либо конкретную программную среду. Она написана тремя весьма уважаемыми прогнозистами, имеющими многолетний опыт работы.

Библиография

Армстронг, Дж. С. (ред.). (2001). Принципы прогнозирования: руководство для исследователей и практиков. Kluwer Academic Publishers. [Amazon, http://buy.geni.us/Proxy.ashx?TSID=140570&GR_URL=http%3A%2F%2Fwww.amazon.com%2Fdp%2F0792379306].

Орд, Дж. К., Филдес, Р. и Курентзес, Н. (2017). Принципы бизнес-прогнозирования. 2-е изд. Wessex Press Publishing Co. [Amazon, http://buy.geni.us/Proxy.ashx?TSID=140570&GR_URL=http%3A%2F%2Fwww.amazon.com%2Fdp%2F0999064916].