

LG 전자 DX (AI + BigData) 교류

5주차 – 2차시

시계열 처리

시계열 데이터 이해와 시계열의 통계모델



소 속 : 경북대학교/인공지능학과

이 름 : 김현철 교수

E-Mail : hyunchul_kim@knu.ac.kr

Timetable

시간표	학습 내용
1차시 08:30 – 09:30	시계열의 개요 및 다루기
2차시 09:40 – 10:50	시계열 데이터를 이용한 회귀 분석 이해
3차시 11:00 – 12:00	시계열 데이터를 이용한 지수 평활

Lunch Break (12:00 – 13:30)

5차시 13:30 - 14:40	시계열 데이터와 회귀 분석 실습
6차시 14:50 – 15:50	시계열 데이터 지수 평활 추정 실습
7차시 16:00 – 16:50	지도교수 미팅
8차시 17:00 – 17:50	지도교수 미팅

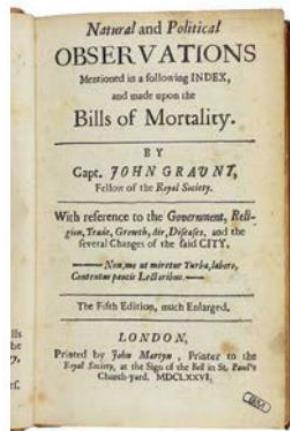
시계열의 개요 및 다루기

• 시계열 분석

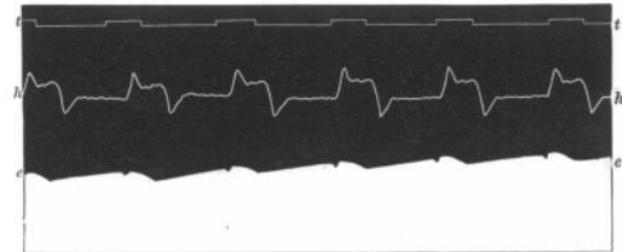
- 시간 순서대로 정렬된 데이터에서 의미 있는 요약과 통계 정보를 추출 하는 방법
- “과거가 미래에 어떤 영향을 주는가?”와 같은 인과관계를 다루는 질문으로 요약

• 다양한 학문 분야에 적용

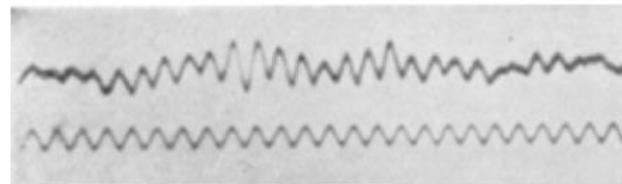
- 의학
- 기상학
- 경제학
- 천문학



시계열식 사고(를 적용한
첫 번째 결과물 중 하나인
존 그란트의 사망표
- 1500년대 초반부터 런
던 교외에 보관되어 있던
사망 기록에 대한 연구를
시작했으며, 이 과정에서
인구통계학을 만듦



1877년에 초기 기록된 ECG



1924년 처음으로 기록된 사람의 EEG

- 일변량(univariate) 시계열
 - 시간에 대해 측정된 변수가 하나만 있는 경우
 - 예, 센서를 통한 시간에 따른 방 온도 측정
- 다변량(multivariate) 시계열
 - 각 타임스탬프에서 측정된 변수가 여러 개인 경우 의미
 - 예, 가속도계를 통해 시간에 따른 x, y, z의 좌표값 측정
- 측정된 여러 변수가 서로 연관되거나, 각 변수가 서로 시간 종속성을 가지는 시계열은 다채로운 분석이 가능

- 시계열 데이터의 일반 문제
 - 누락된 데이터 (결측 데이터)
 - 시계열의 빈도 변경 (업샘플링, 다운샘플링)

- 누락된 데이터 다루기

- 의료 분야를 예로 들면...

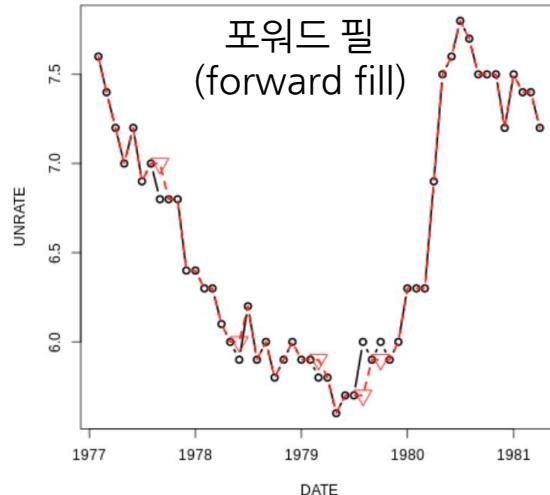
- 환자가 필요한 진료를 받지 않는 경우
 - 환자의 건강 상태가 좋아서 추가 진료가 필요 없는 경우
 - 환자를 잊었거나 치료가 불충분했던 경우
 - 의학 장치가 제멋대로 기술적인 오작동을 일으킨 경우
 - 데이터 입력 시 오류가 발생한 경우

- 해결하는 가장 일반적인 방법

- 대치법(imputation): 데이터셋 전체의 관측에 기반하여 누락된 데이터를 채워 넣는 방법
 - 보간법(interpolation): 대치법의 한 형태로 인접한 데이터를 사용하여 누락된 데이터를 추정하는 방법
 - 영향 받은 기간 삭제: 누락된 데이터의 기간을 완전히 사용하지 않는 방법

1 시계열 데이터 정리

- 대치법(imputation)
 - 포워드 필(forward fill)
 - 누락된 값이 나타나기 직전의 값으로 누락된 값을 채우는 방법
 - 어떤 수학적인 원리나 복잡한 논리가 필요 없음
 - 백워드 필(backward fill)
 - 과거의 값으로 이후에 발생한 채울 수 있듯이 이전 값을 채우는 것도 가능
 - 단, 이는 사전 관찰이기 때문에 데이터를 사용하여 미래를 예측하지 않거나 특정 분야의 지식에 기반하여 데이터의 미래보다 과거를 채우는 게 더 의미 있는 경우에만 사용



실선은 원본, 점선은 임의로 누락된 값을 포워드
필로 채운 시계열을 나타내며, 역 삼각형 기호는
포워드 필로 채워진 값의 부분을 의미

사전관찰이란?

시계열 분석에서 **사전관찰**이란 용어는 미래의 어떤 사실을 안다는 뜻으로 사용됩니다. 이러한 지식은 모델의 설계, 학습, 검증 단계에서는 알 수 없습니다. 사전관찰은 데이터를 통해 실제로 알아야 하는 시점보다 더 일찍 미래에 대한 사실을 발견하는 방법입니다.

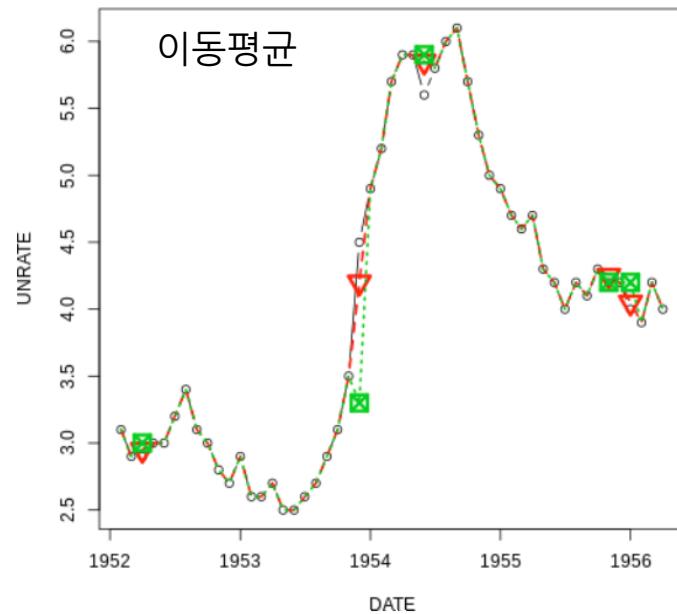
사전관찰은 어떻게든 미래에 일어날 일에 대한 정보가 모델에서 시간을 거슬러 전파되어 모델의 초기 동작에 영향을 주는 방법입니다. 예를 들어 모델에 대한 하이퍼파라미터를 선택할 때 데이터셋에서 다양한 시간에 모델을 시험한 다음 최적의 모델을 하나 고르는 겁니다. 그러면 해당 모델의 테스트를 데이터 시작 부분에서 다시 할 수 있습니다. 이때 한 시점에 대해 다음에 일어날 일을 아는 모델을 선택했기 때문에 문제가 있다고 볼 수 있습니다.

아쉽게도 사전관찰을 위한 자동화된 코드나 통계적 테스트가 아직 없어서 항상 고민해야 합니다.

- 대치법(imputation)

- 이동평균(moving average)

- 롤링 평균 또는 중앙값으로 데이터를 대치할 수 있음
 - 과거의 값으로 미래의 값을 “예측” 한다는 관점에서는 포워드 필과 유사
 - 단, 이동평균은 최근 과거의 여러 시간대를 입력한 내용을 사용한다는 점에서 다름

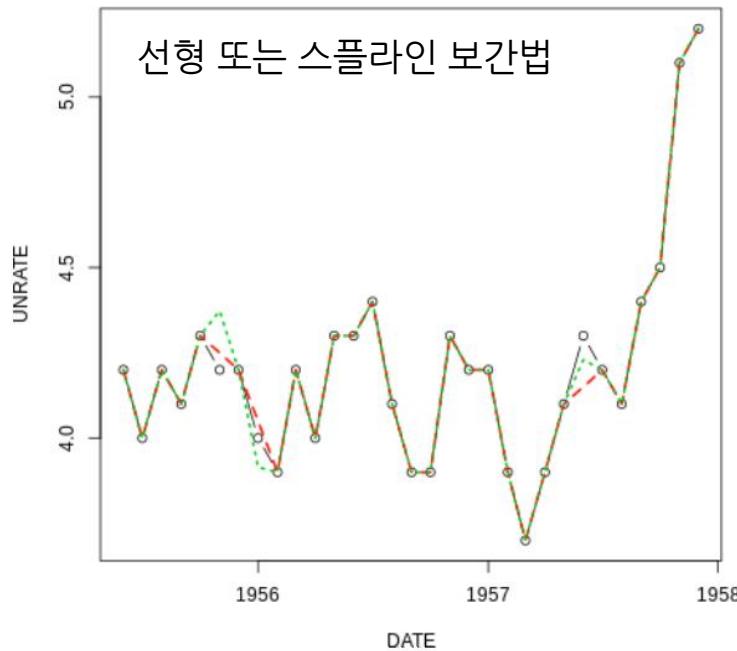


점선은 사전관찰이 없는 이동평균 대치를, 파선은 사전관찰이 있는 이동평균 대치를 보여줍니다. 마찬가지로, 사각형은 사전관찰 없는 이동평균을 대치한 값, 역삼각형은 사전관찰이 있는 이동평균을 대치한 값을 의미

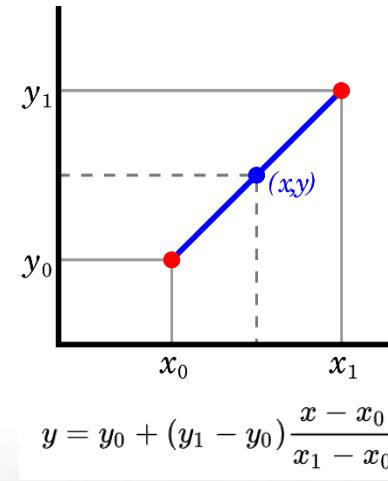
- 보간법(interpolation):

- 선형 또는 스플라인 보간법(linear or spline interpolation)

- 누락된 데이터가 주변 데이터에 선형적 또는 비선형적 일관성을 갖도록 제한
- 과거와 미래의 데이터를 모두 활용하거나 둘 중 하나만 활용 할 수 있음



파선은 선형 보간법을, 점선은 스플라인 보간법을 보여줍니다.



1 시계열 데이터 정리

- 데이터가 누락될 만한 잠재적 이유와 수정사항의 잠재적 영양 등을 신중하게 고려
 - 데이터가 무작위로 누락되었다는 사실을 증명하는 것은 실제로 불가능합니다. 하지만 현실 세계에서 발생하는 대부분의 누락된 값이 무작위로 생겨났을 가능성은 낮습니다.
 - 측정된 변수를 통해 누락이 발생할 확률을 설명할 수도 있지만, 그렇지 않은 경우도 있습니다. 누락된 데이터의 패턴을 설명하는 가장 좋은 방법은 많은 특징 열을 가진 데이터를 조사하는 것입니다. 하지만 표준적인 시계열 분석 방법은 아닙니다.
 - 누락된 데이터를 대치한 값이 만들어 내는 불확실성에 대한 이해가 필요하다면, 다양한 시나리오를 수행하거나 가능한 한 데이터 수집 과정에 연관된 여러 사람과 대화를 나눠봐야 합니다.
 - 누락된 데이터의 처리 방법은 이어지는 작업의 데이터 용도를 설명할 수 있어야 합니다. 신중하게 사전관찰이 생기지 않도록 조심해야 하고, 후속 작업의 유효성에 사전관찰이 얼마나 심각한 영향을 미칠지에 대한 판단을 내려야 합니다.

1 시계열 데이터 정리

- 업샘플링(upsampling)과 다운샘플링(downsampling)
 - 시계열 데이터의 출처가 서로 다르면 샘플링 빈도가 같은 경우가 많음.
 - 이러한 상황이 데이터의 샘플링 빈도를 바꿔야하는 경우가 생김
 - 타임스탬프의 빈도를 늘이거나 줄이는 방법을 업샘플링과 다운샘플링이라고 함

NOTE _ 다운샘플링은 원본 시계열보다 타임스탬프가 더 낮은 빈도로 발생하게끔 데이터의 부분집합을 만듭니다. 업샘플링은 데이터가 실제보다 더 자주 수집된 것처럼 데이터를 표현합니다.

- 시계열 시각화

- 데이터 분석 작업에서 가장 먼저 해야하는 것은 데이터를 그래프으로 나타내는 것입니다. 그래프는 패턴, 특이한 관측값, 시간에 따른 변화, 변수 사이의 관계 등의 데이터의 많은 특징을 눈으로 볼 수 있게 해줍니다.
- 데이터를 그림으로 나타낸 그래프에서 보이는 특징은 사용할 예측 기법에 반드시 포함되어야 합니다. 데이터의 종류가 어떤 예측 기법을 사용할 지 결정하기도 하고, 어떤 그래프가 적절한지도 알려 줍니다. 하지만 그래프를 그리기 전에, R에서 그릴 시계열을 준비해야 합니다.

2.1 ts 객체

- **ts 객체**

- 시계열이란 각 숫자가 기록된 시간에 관한 정보가 있는 숫자들의 목록으로 생각할 수 있습니다. R에서는 이러한 정보를 ts 객체로 저장할 수 있습니다.

연도	관측값
2012	123
2013	39
2014	78
2015	52
2016	110

- `ts()` 함수로 이것을 ts 객체로 바꾸겠습니다:

```
y <- ts(c(123,39,78,52,110), start=2012)
```

- 1년에 한 관측값이 있는 연간 데이터를 가지고 있다고 가정했으니, 시작 연도만(또는 마지막 연도만) 있으면 됩니다.

2.2 시간 그래프

- 시간 그래프
 - 시계열 데이터에서, 가장 먼저 그려야 할 것은 시간 그래프(time plot)입니다. 즉, 관측값을 관측 시간에 따라 인접한 관측값을 직선으로 연결하여 그리는 것입니다.
 - 아래의 그림은 호주에서 가장 큰 두 도시 사이를 운항하는 Ansett 항공의 이코노미석 주별 수송량을 나타냅니다.

```
autoplot(melsyd[, "Economy.Class"]) +  
  ggtitle("이코노미석 탑승객: 멜버른-시드니") +  
  xlab("연도") +  
  ylab("탑승객(단위: 1000명)")
```



Ansett 항공 주별 이코노미석 탑승객

2.3 시계열 패턴

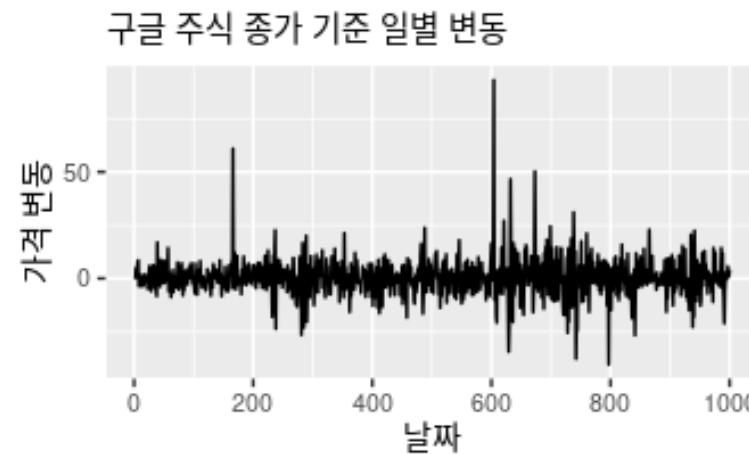
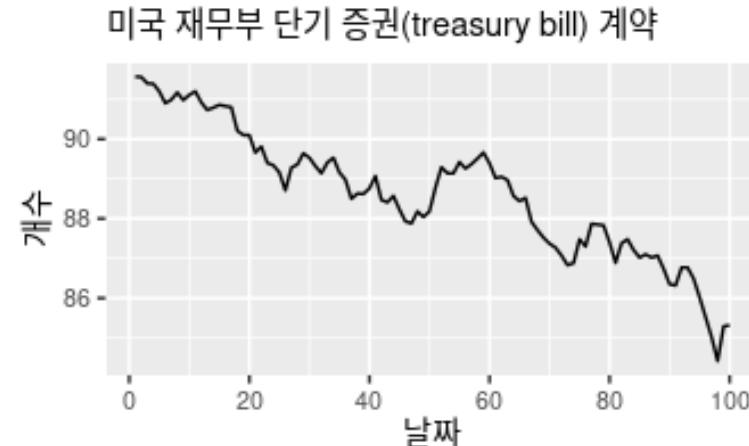
- 시계열 패턴
 - 시계열을 설명하려면, 지금까지 사용한 “추세(trend)”와 “계절성(seasonality)” 같은 단어를 좀 더 조심스럽게 정의해야 합니다.
 - 추세(trend)
 - 데이터가 장기적으로 증가하거나 감소할 때, 추세(trend)가 존재합니다. 추세가 선형적일 필요는 없습니다. 때때로 어떤 추세가 증가에서 감소로 변화하는 경우에, 그것을 추세의 “방향이 변화했다”라고 언급할 것입니다.
 - 계절성(seasonality)
 - 해마다 어떤 특정한 때나 1주일마다 특정 요일에 나타나는 것 같은 계절성 요인이 시계열에 영향을 줄 때 계절성(seasonality) 패턴이 나타납니다. 계절성은 빈도의 형태로 나타나는데, 그 빈도는 항상 일정하며 알려져 있습니다. 위의 당뇨병 약 월별 매출액에는 계절성이 나타나는데, 이 계절성은 부분적으로 연말에 발생하는 약품 가격 변동에 의한 것입니다.
 - 주기성(cycle)
 - 고정된 빈도가 아닌 형태로 증가나 감소하는 모습을 보일 때 주기(cycle)가 나타납니다. 보통 이러한 요동은 경제 상황 때문에 일어나고, 흔히 “경기 순환(business cycle)”과 관련 있습니다. 보통 이러한 요동의 지속 기간은 적어도 2년 이상입니다.

2.3 시계열 패턴

- 시계열 패턴
 - 많은 이들이 주기적인 패턴과 계절적인 패턴을 혼동하지만, 사실들은 정말 다릅니다.
 - 일정한 빈도로 나타나지 않는 요동은 주기적입니다. 빈도가 변하지 않고 연중 어떤 시기와 연관되어 있다면 그 요동은 계절성입니다.
 - 일반적으로, 주기들의 평균 길이는 계절성 패턴의 길이보다 길고, 주기의 크기는 계절적인 패턴의 크기보다 좀 더 변동성이 큰 경향이 있습니다.
 - 많은 시계열에는 추세(trend), 계절성(seasonality), 주기(cycle)가 있습니다. 예측 기법을 고를 때, 먼저 데이터에서 나타나는 시계열 패턴을 살펴봐야할 것이고, 그 다음 적절하게 패턴을 잡아낼 수 있는 기법을 선택해야할 것입니다.

2.3 시계열 패턴

• 시계열 패턴



서로 다른 패턴을 나타내는 4가지 시계열 예제.

2.3 시계열 패턴

- 시계열 패턴

- 월별 주택 매매 그래프에서(왼쪽 위) 매년 강한 계절성과 약 6–10년의 몇몇 강한 주기적인 패턴이 보입니다. 전체 기간에 걸쳐 데이터에 분명한 추세가 있지는 않습니다.

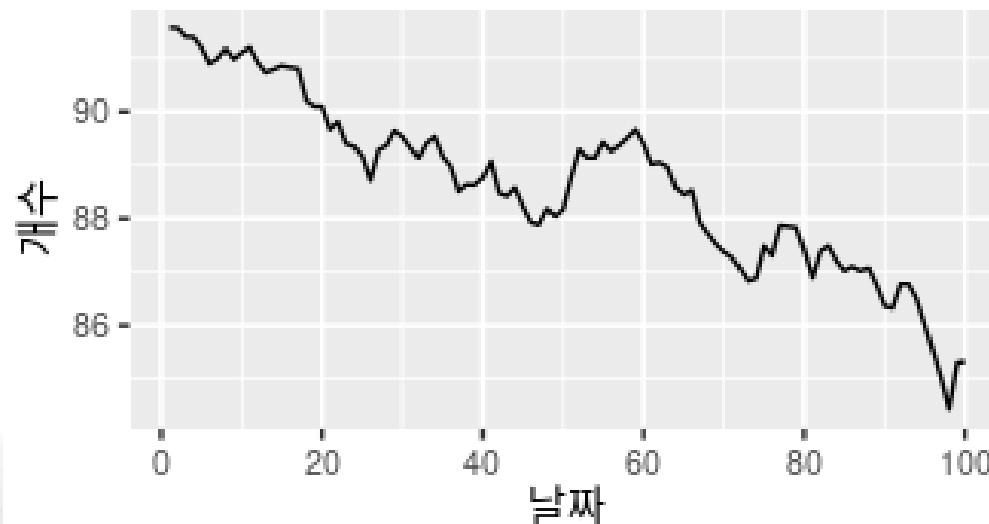


2.3 시계열 패턴

- 시계열 패턴

- 미국 재무부 단기 증권 그래프는(오른쪽 위) 1981년 시카고 시장의 100 연속 거래일 데이터를 나타냅니다. 여기에는 계절성은 없지만, 아래로 내려가는 추세가 분명하게 있습니다. 아마도 훨씬 더 긴 시계열이 있었다면, 이 하향 추세가 실제로는 긴 주기의 한 부분이라는 것을 볼 수도 있었을 것이지만, 100 거래일만 놓고 볼 때, 추세가 있는 것 같습니다.

미국 재무부 단기 증권(treasury bill) 계약

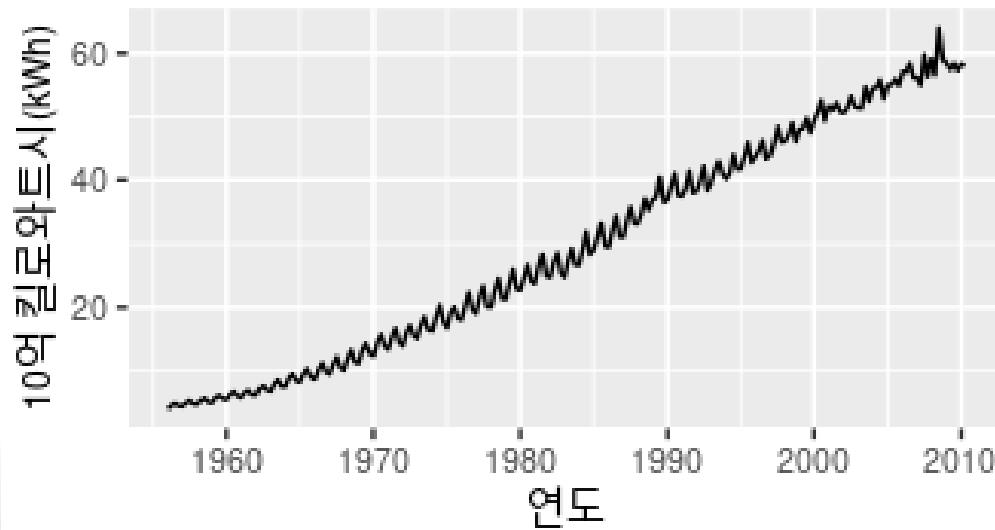


2.3 시계열 패턴

• 시계열 패턴

- 호주 분기별 전기 생산량 그래프는(왼쪽 아래) 강한 계절성과 함께 강한 증가 추세를 나타냅니다. 여기에서 주기적인 행동이 보이지는 않습니다.

호주 분기별 전력 생산

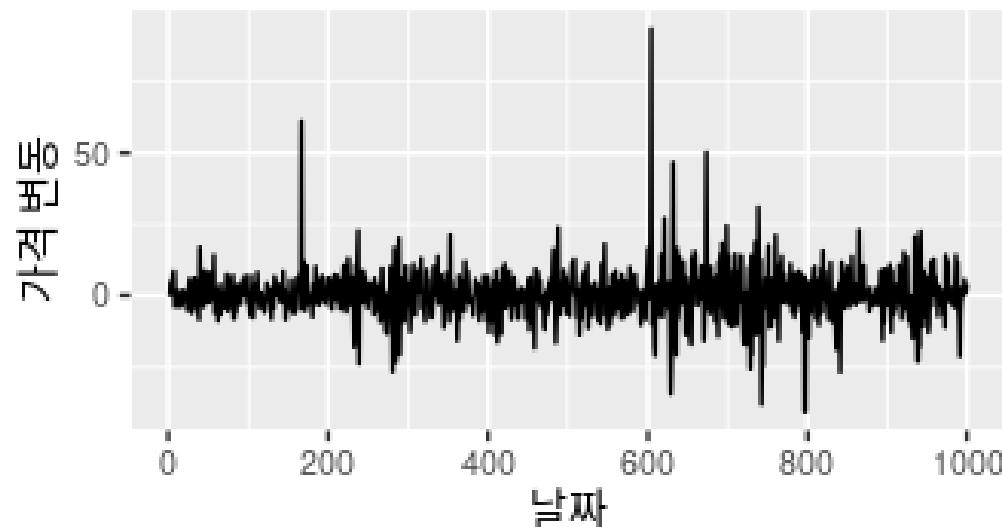


2.3 시계열 패턴

• 시계열 패턴

- 구글(Google) 주식의 일별 종가(거래일 마지막 가격) 그래프(오른쪽 아래)에는 추세나 계절성 또는 주기적인 행동이 없습니다. 잘 예측할 수 있을 것 같지 않은 무작위적인 요동이 있고, 예측 모델을 만드는데 도움이 될 강한 패턴도 없습니다

구글 주식 종가 기준 일별 변동

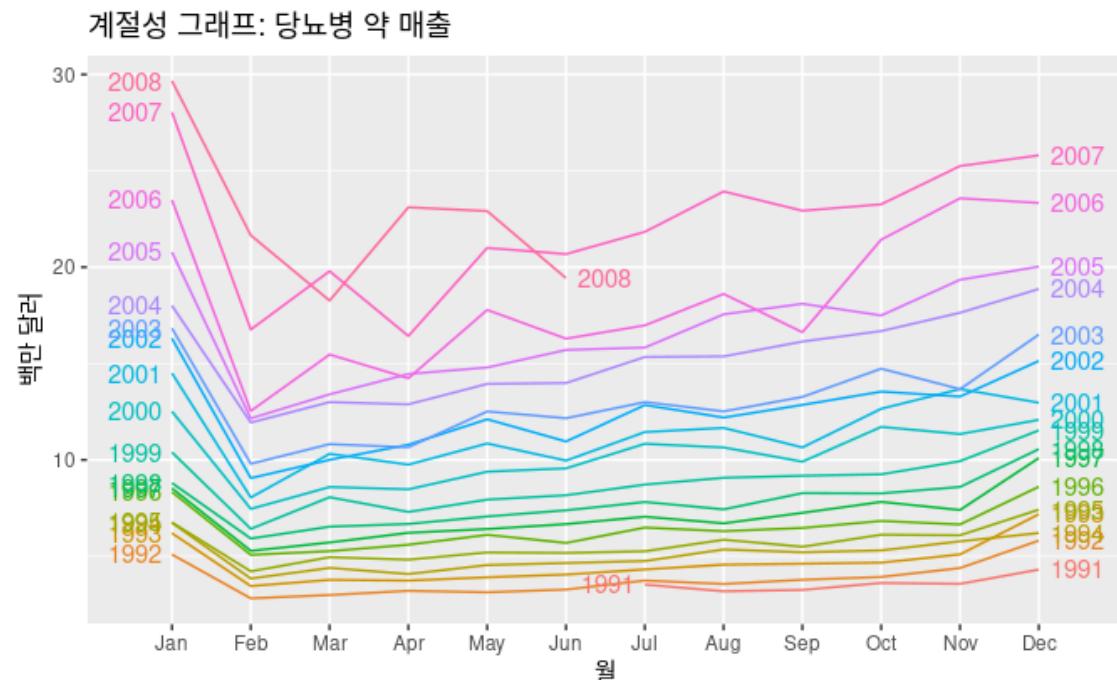


2.4 계절성 그래프

- 계절성 그래프
 - 계절성 그래프(seasonal plot)는 각 “계절(season)”에 대해 관측한 데이터를 나타낸다는 점만 제외하고는 시간 그래프와 비슷합니다. 아래 예제에서는 당뇨병 약 매출을 나타냅니다.

```
ggseasonplot(a10, year.labels = TRUE, year.labels.left = TRUE) +  
  ylab("백만 달러") +  
  xlab("월") +  
  ggtitle("계절성 그래프: 당뇨병 약 매출")
```

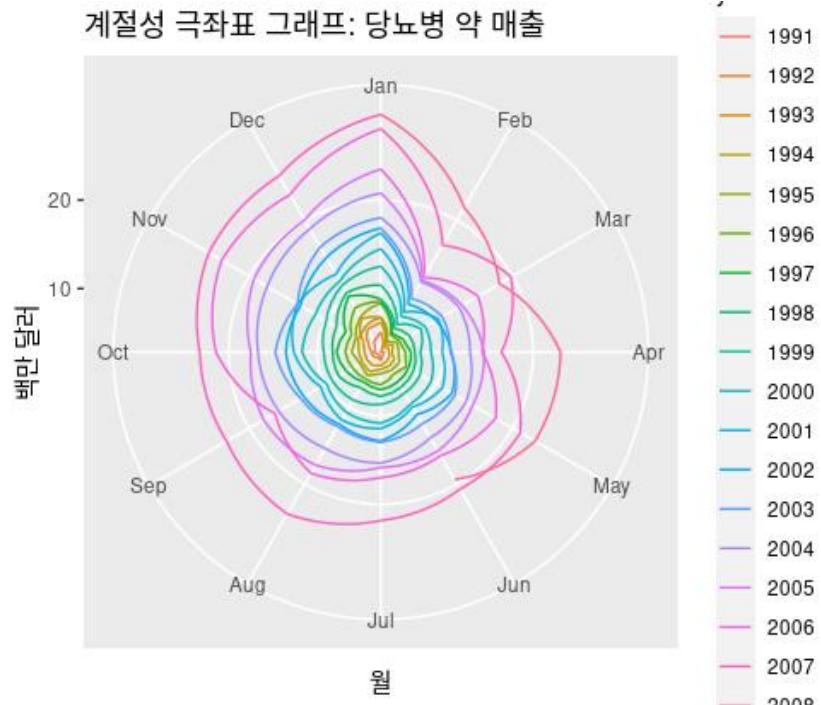
- 이 경우에, 분명히 매년 1월에 매출이 크게 뛩니다. 실제로는, 소비자들이 연말에 사재기해서 나타나는 매출이 1~2주 이후까지 정부에 등록되지 않아서 나타나는 현상일 것입니다.
- 이 그래프는 2008년 3월 매출이 비정상적으로 작았다는 것도 보여줍니다(다른 연도에서는 2월과 3월 사이에 증가합니다). 2008년 6월 매출이 작은 이유는 당시에 데이터를 모을 때 제대로 집계하지 않았기 때문일 것입니다.



2.4 계절성 그래프

- 계절성 그래프
 - 극좌표(polar coordinate)를 사용하여 계절성 그래프를 바꿔 나타내면 도움이 됩니다. 아래에서 볼 수 있는 것처럼, polar=TRUE로 두면, 수평축 대신 원형축으로 나타냅니다

```
ggseasonplot(a10, polar = TRUE) +  
  ylab("백만 달러") +  
  xlab("월") +  
  ggtitle("계절성 극좌표 그래프: 당뇨병 약 매출")
```



호주 당뇨병 약 월별 매출의 계절성 극좌표 그래프.

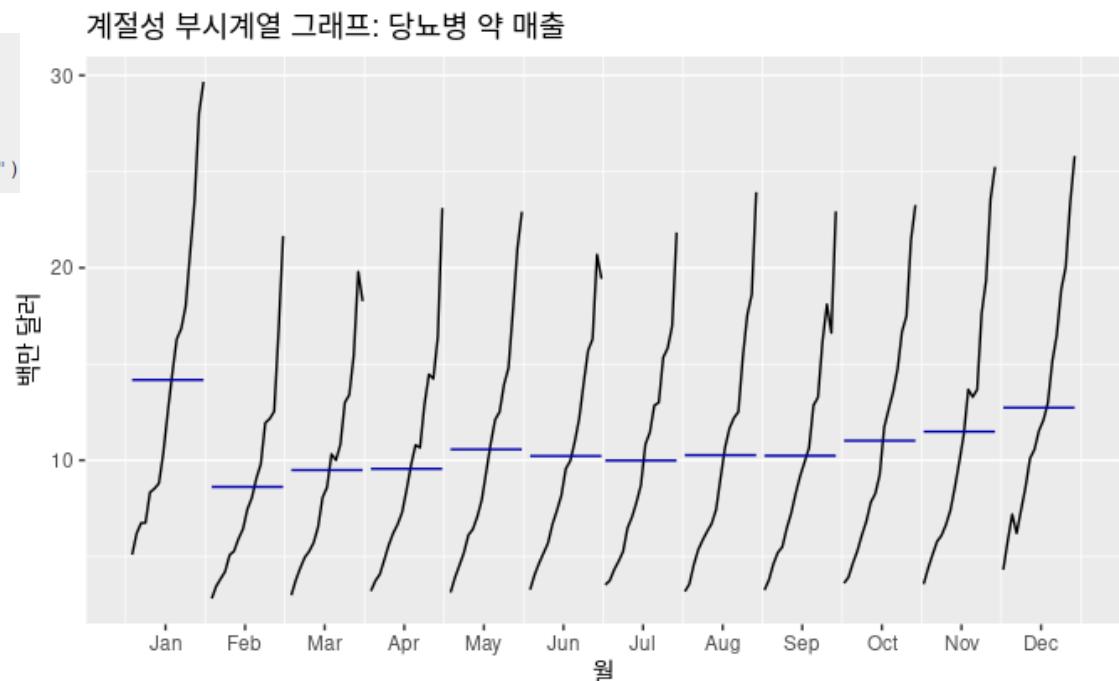
2.5 계절성 부시계열 그래프

• 계절성 부시계열 그래프

- 계절성 패턴을 강조하여 나타내는 또 다른 방법은 각 계절에 대한 데이터를 모아서 분리된 작은 시간 그래프로 나타내는 것입니다.

```
ggsubseriesplot(a10) +  
  ylab("백만 달러") +  
  xlab("월") +  
  ggtitle("계절성 부시계열 그래프: 당뇨병 약 매출")
```

- 수평선은 각 월에 대한 평균값을 의미합니다. 이러한 형태의 그래프로 중요한 계절성 패턴을 분명하게 살펴볼 수 있고, 계절성이 시간에 따라 어떻게 변하는지 도 볼 수 있습니다.
- 특정한 철에서 나타나는 변화를 확인할 때 특별히 쓸모가 있습니다. 이 예제에서, 이 그래프가 특별한 것을 보여주지는 않습니다만, 몇몇 경우에서 시간에 따른 계절성 변화를 나타낼 때 가장 쓸모 있는 방법입니다.



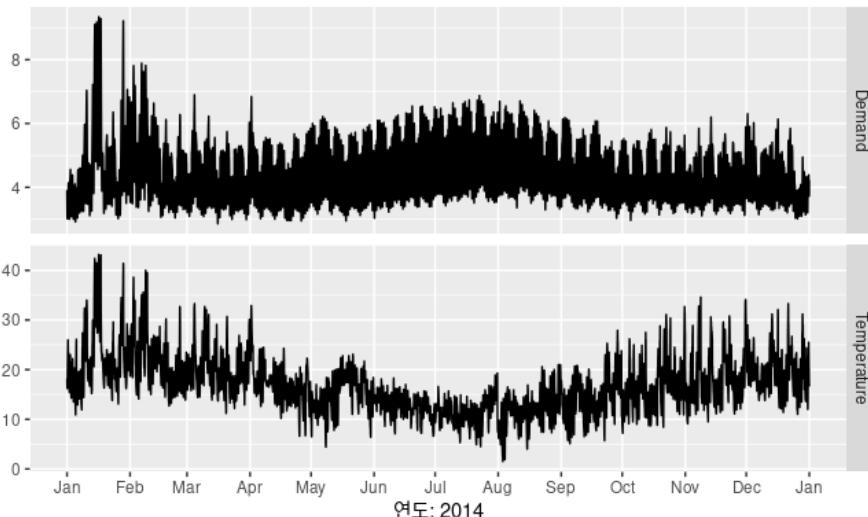
호주 당뇨병 약 월별의 계절성 부시계열 그래프.

2.6 산점도

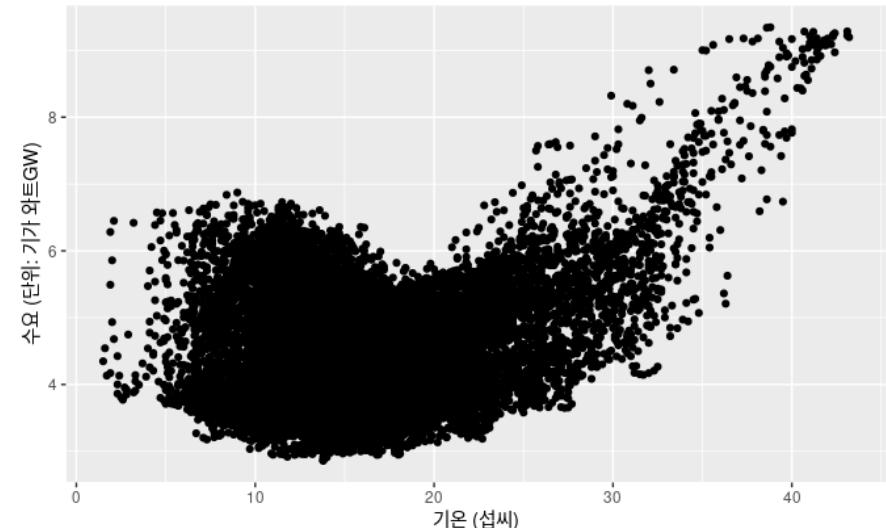
- 산점도
 - 지금까지 다룬 그래프는 각각의 시계열을 시각화 할 때 유용합니다. 시계열 사이의 관계를 살필 때도 쓸모가 있습니다.

```
qplot(Temperature, Demand, data=as.data.frame(elecemand)) +  
  ylab("수요 (단위: 기가 와트GW)") + xlab("기온 (섭씨)")
```

호주 빅토리아 주 30분 단위 전력 수요



2014년 호주 빅토리아 주 30분 단위 전력 수요와 기온.



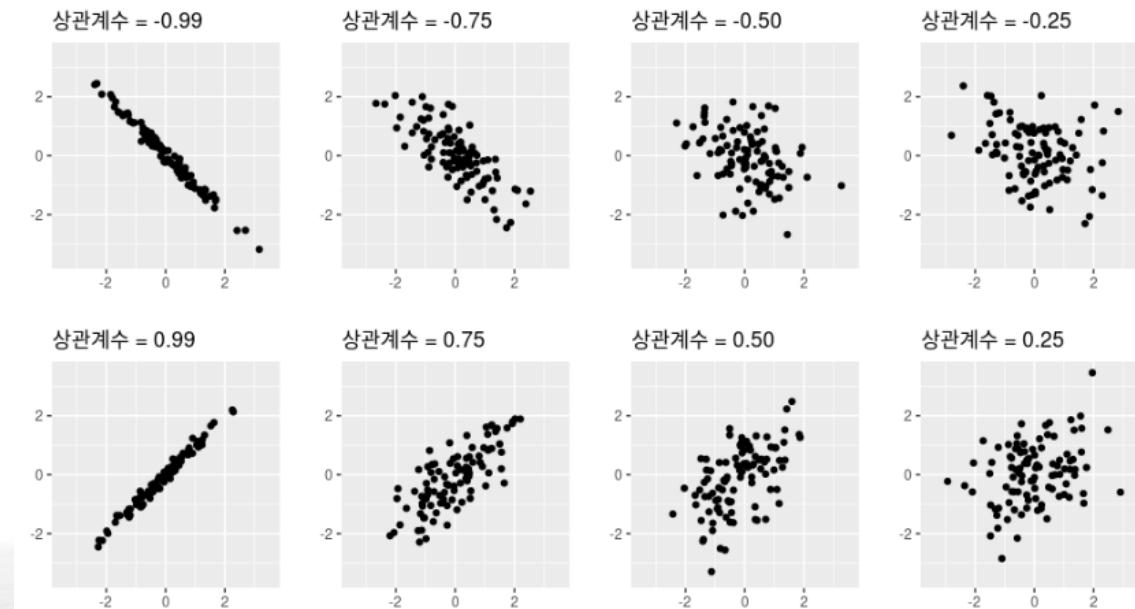
2.6 산점도

-

상관관계

- 상관계수(correlation coefficient)는 두 변수 사이의 관계의 강도를 측정할 때 흔히 계산하는 양입니다. 두 변수 x와 y 사이의 상관계수는 다음과 같이 주어집니다.

$$r = \frac{\sum(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum(x_t - \bar{x})^2} \sqrt{\sum(y_t - \bar{y})^2}}.$$

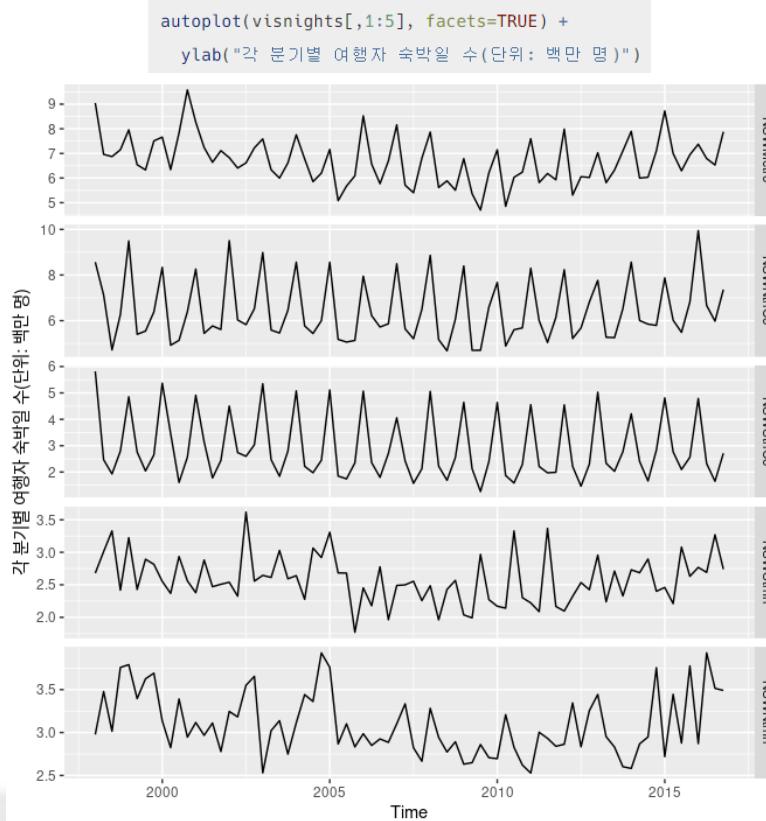


상관계수가 다른 데이터를 나타내는 예제.

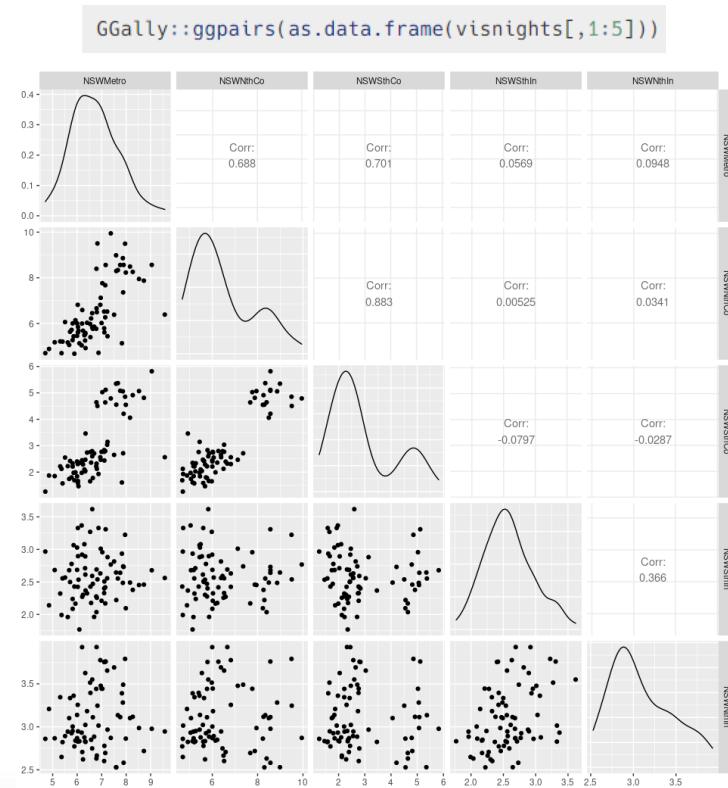
2.6

산점도

- 산점도행렬
 - 몇 가지 후보가 될 만한 예측변수(predictor variable)가 있을 때, 각 변수를 다른 변수에 대해 나타내는 것이 도움이 됩니다.



호주 NSW의 여러 지역에서 분기별 여행자 숙박일 수.

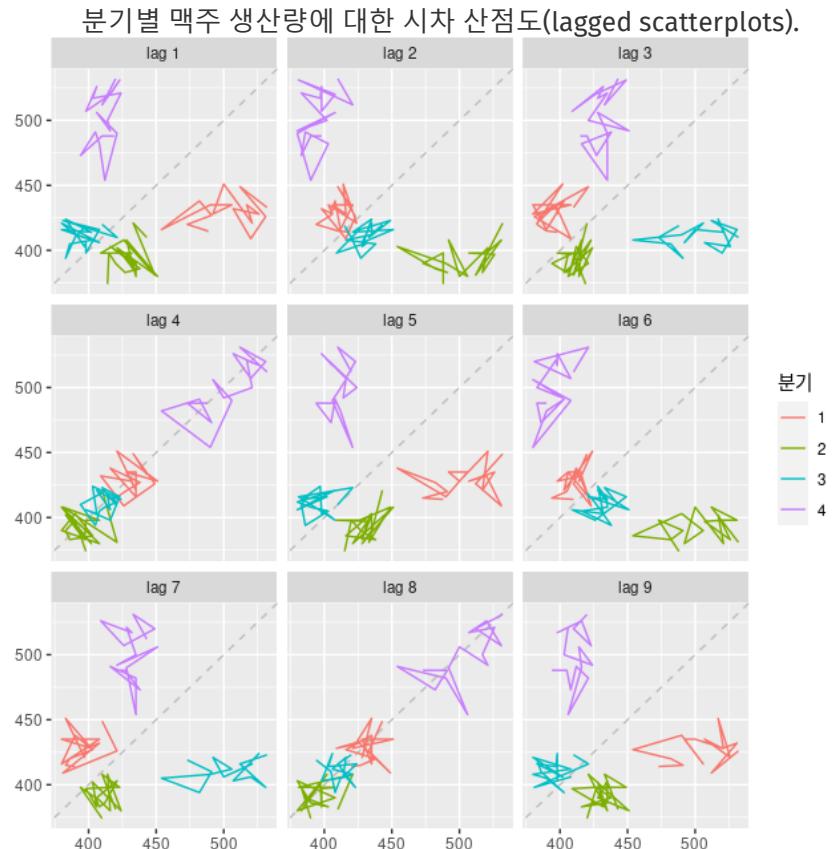


호주 NSW의 5개 지역에서 분기별 여행자 숙박일 수의 산점도행렬.

2.7 시차 그래프

• 시차 그래프

- 분기별 호주 맥주 생산량의 산점도(scatterplot)를 나타냅니다.
 - 여기에서 가로축은 시계열의 시차값(lagged value)을 나타냅니다. 각 그래프는 서로 다른 k 값에 대해서 y_t (세로축)를 y_{t-k} (가로축)에 대해 나타낸 것입니다.



```
beer2 <- window(ausbeer, start=1992)
gglagplot(beer2) +
  guides(colour=guide_legend(title="분기")) +
  ggtitle("")
```

- 세로축(y_t)의 범례에 있는 색상은 분기별 변수를 나타냅니다. 시차 4와 시차 8에서 양의 관계가 뚜렷합니다. 이는 데이터에 계절성이 강하게 있다는 것을 반영합니다.
- 저점(2분기)에 대해 최고점(4분기)을 나타냈기 때문에 시차 2와 시차 6에서 음의 관계가 나타납니다

2.8 자기상관

- 자기상관
 - 상관값이 두 변수 사이의 선형 관계의 크기를 측정하는 것처럼, 자기상관(autocorrelation)은 시계열의 시차 값(lagged values) 사이의 선형 관계를 측정합니다.
 - 시차 그래프(lag plot)에서 각 패널과 관련된 몇 가지 자기상관 계수가 있습니다. 예를 들면, r_1 은 y_t 와 y_{t-1} 사이의 관계를 측정하고, r_2 는 y_t 와 y_{t-2} 사이의 관계를 측정하는 식입니다.

$$r_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2},$$

여기에서 T 는 시계열의 길이입니다.

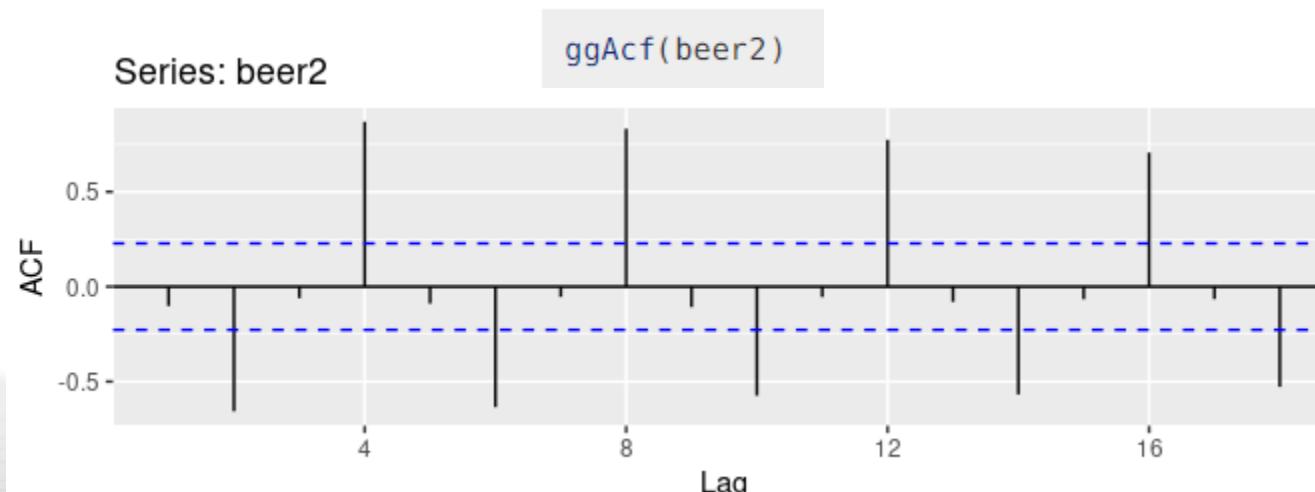
2.8 자기상관

- 자기상관

- 다음의 표는 맥주 생산량 데이터에 대한 처음 9개의 자기상관 (autocorrelation) 계수를 보여줍니다.

r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9
-0.102	-0.657	-0.060	0.869	-0.089	-0.635	-0.054	0.832	-0.108

- 이 값은 “호주 맥주 생산량의 산점도” 그림에서의 9개의 산점도 (scatterplot)와 대응됩니다. 자기상관 계수는 보통 자기상관함수 (ACF)를 나타내기 위해 그립니다. 이 그래프는 상관도표 (correlogram)라고도 알려져 있습니다.



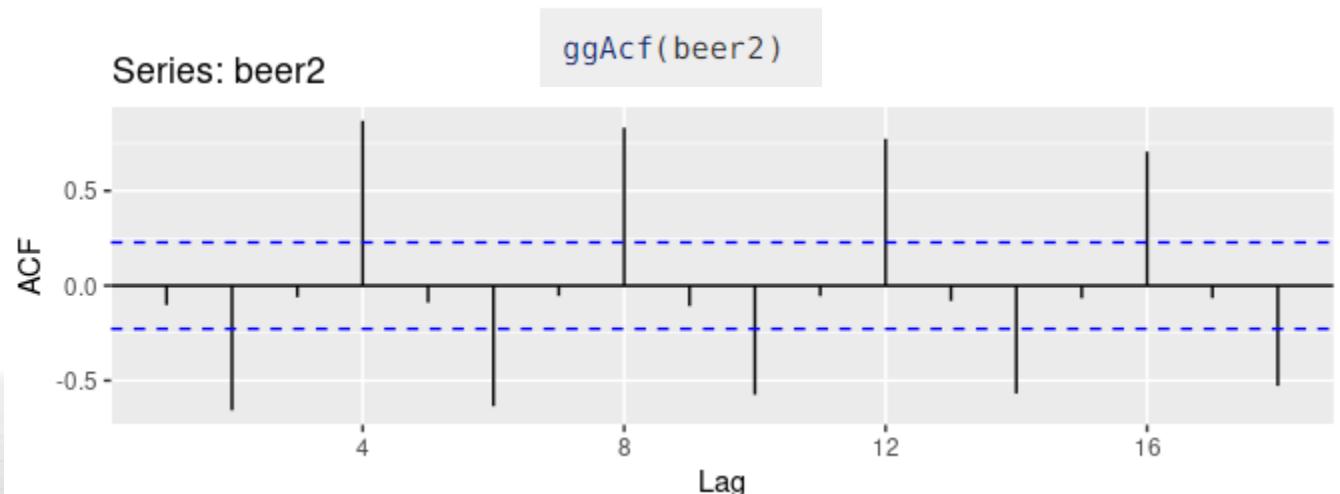
분기별 맥주 생산량에 대한 자기상관함수.

2.8 자기상관

- 자기상관

- 이 그래프에서:

- r_4 는 다른 시차들보다 값이 높은데, 이것은 데이터의 계절성 패턴 때문입니다:
 - 고점은 4개의 분기마다 나타나는 경향이 있고, 마찬가지로 저점 역시 4 개의 분기마다 나타나는 경향이 있습니다.
 - r_2 는 다른 시차들보다 더 큰 음의 값을 나타내는데, 왜냐하면 저점이 고점 직후의 2개 분기마다 나타나는 경향 때문입니다.
 - 파란 점선은 상관계수가 0과 유의하게 다른지 아닌지를 나타냅니다.



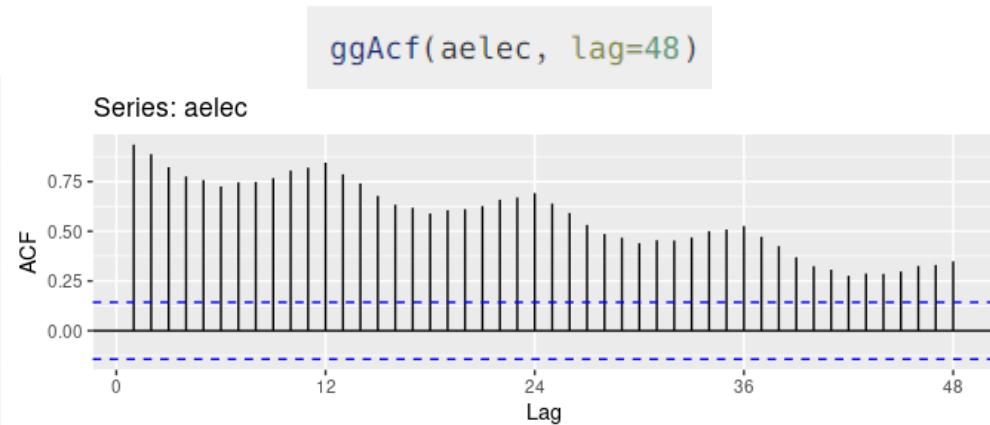
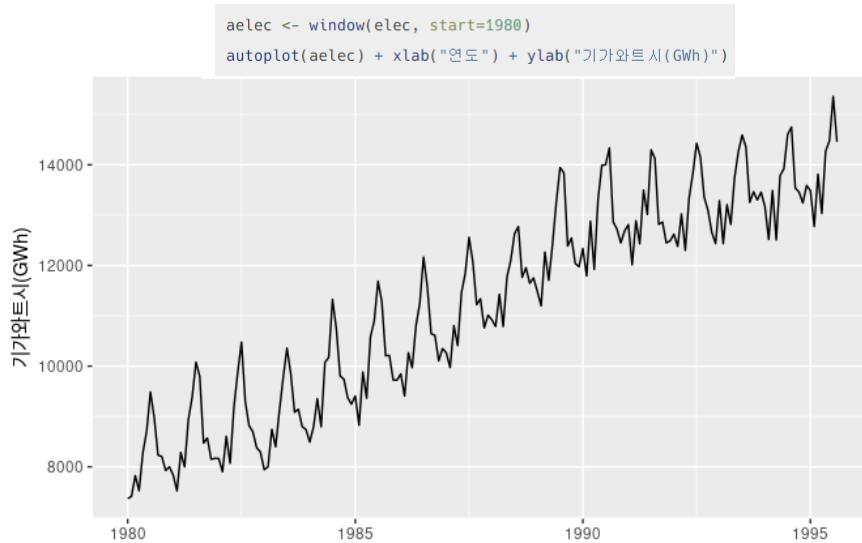
분기별 맥주 생산량에 대한 자기상관함수.

2.8 자기상관

- ACF 그래프에서 추세와 계절성
 - 데이터에 추세가 존재할 때, 작은 크기의 시차에 대한 자기상관은 큰 양의 값을 갖는 경향이 있는데, 왜냐하면 시간적으로 가까운 관측치들이 관측값의 크기에 있어서도 비슷하기 때문입니다.
 - 그래서 추세가 있는 시계열의 ACF는 양의 값을 갖는 경향이 보이며, 이러한 ACF의 값은 시차가 증가함에 따라 서서히 감소한다.
 - 데이터에 계절성이 존재할 때의 자기상관은 다른 시차의 경우보다 계절성 시차(계절성 빈도의 배수로 나타나는)의 경우에 더 크게 나타날 것입니다.

2.8 자기상관

- ACF 그래프에서 추세와 계절성
 - 추세와 계절성을 모두를 나타내는 데이터의 경우에는 이들의 조합된 효과를 확인할 수 있습니다. 왼쪽그림의 호주 월별 전력 수요 시계열은 추세와 계절성 양쪽 모두를 나타냅니다. 이 시계열의 ACF은 오른쪽 그림에 있습니다.



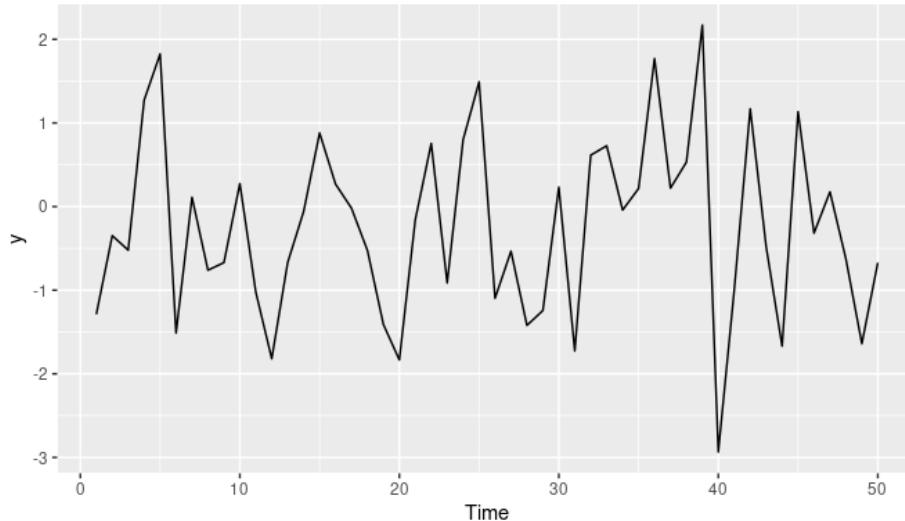
월별 호주 전력 수요의 ACF.

- 시차값(lags)이 증가할수록 ACF 값이 서서히 감소하는 것은 추세 때문인 반면, “물결 모양(scalloped)”이 나타나는 것은 계절성 때문입니다.

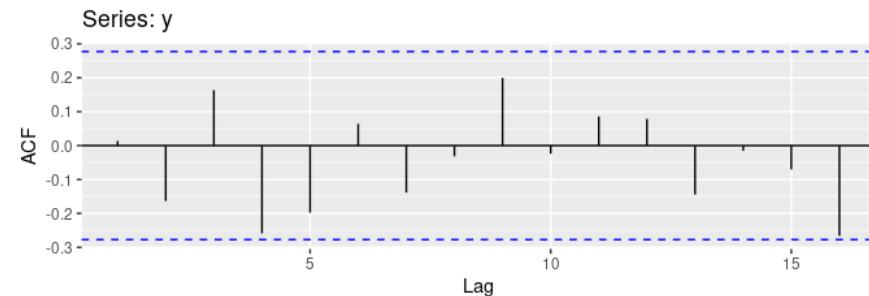
- 백색잡음

- 자기상관(autocorrelation)이 없는 시계열을 백색잡음(white noise)이라고 부릅니다. 아래 그림은 백색잡음 시계열의 예를 보여줍니다.

백색잡음



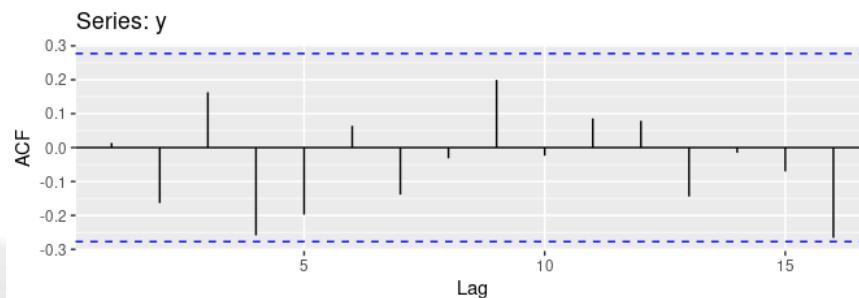
백색잡음 시계열.



백색잡음 시계열에 대한 자기상관함수.

- 백색잡음

- 백색잡음 시계열에 대해, 자기상관 값 각각이 거의 0일 것이라고 기대할 수 있습니다. 물론, 어떤 무작위적 변동 때문에 자기상관값이 정확하게 0은 아닐 것입니다.
- 백색잡음 시계열에 대해, ACF에서 뾰족한 막대의 95% 가 $\pm 2\sqrt{T}$ 에 들어갈 것이라고 기대할 수 있습니다. 여기에서 T 는 시계열의 길이입니다.
- ACF의 그래프에서 이러한 경계를 주로 그래프로 그려서 나타냅니다(위의 파란 점선 경계). 하나 이상의 커다란 뾰족한 막대가 이러한 경계를 벗어나거나 뾰족한 막대의 5% 이상이 이러한 경계를 벗어난다면, 이러한 시계열은 아마도 백색잡음이 아닐 것입니다.



백색잡음 시계열에 대한 자기상관함수.

- 예제에서는 $T=50$ 이라서 $\pm 2\sqrt{50} = \pm 0.28$ 에 경계가 있습니다. 자기상관 계수 전부가 이러한 한도 안에 들어가는 모습이 데이터가 백색잡음이라는 사실을 말해줍니다.

3 예측가의 도구 상자

- 예측가의 도구 상자

- 이 장에서는, 다양한 예측 상황에서 사용할 수 있는 일반적이면서 좋은 도구 몇 가지를 다룹니다.
- 변환과 조정으로 예측 작업을 단순하게 만드는 방법, 예측 기법에서 이용 가능한 정보를 적절하게 사용했는지 확인하는 기법, 그리고 예측구간(prediction interval)을 계산하는 기법 등의 예측 기법을 벤치마크하는 방법을 살펴볼 것입니다.

3.1 몇 가지 단순한 예측 기법

- 몇 가지 단순한 예측 기법

- 몇 가지 예측 기법은 아주 단순하고 놀라울만큼 효과적입니다. 다음의 4가지 예측 기법을 이 책에서 벤치마크할 때 사용하겠습니다.

- 평균 기법(Average method)
 - 단순 기법(naïve method)
 - 계절성 단순 기법(Seasonal naïve method)
 - 표류 기법(drift)

3.1 몇 가지 단순한 예측 기법

- 평균 기법(Average method)
 - 이 경우에는, 예측한 모든 미래 값은 과거 데이터의 평균과 같습니다. 과거 데이터를 y_1, \dots, y_T 라고 쓴다면, 예측값을 다음과 같이 쓸 수 있습니다.

$$\hat{y}_{T+h|T} = \bar{y} = (y_1 + \dots + y_T)/T.$$

- $\hat{y}_{T+h|T}$ 은 데이터 y_1, \dots, y_T 로 y_{T+h} 를 추정한 것을 단순하게 쓴 것입니다.

```
meanf(y, h)
# y에는 시계열이 들어갑니다
# h에는 예측 범위가 들어갑니다
```

3.1 몇 가지 단순한 예측 기법

- 단순 기법
 - 단순 기법(naïve method)에서는 모든 예측값을 단순하게 마지막 값으로 둡니다. 즉,

$$\hat{y}_{T+h|T} = y_T.$$

- 이 기법은 다양한 경제 금융 시계열을 다룰 때 상당히 잘 들어맞습니다.

```
naive(y, h)
```

```
rwf(y, h) # 위의 것과 같은 역할을 하는 함수
```

- 데이터가 확률보행(random walk) 패턴을 따를 때는 단순 기법(naïve method)이 최적이라서, 확률보행 예측값(random walk forecasts)이라고 부르기도 합니다.

3.1 몇 가지 단순한 예측 기법

- 계절성 단순 기법(Seasonal naïve method)
 - 계절성이 아주 뚜렷한 데이터를 다룰 때 비슷한 기법이 유용합니다. 이 경우에는, 각 예측값을 연도의 같은 계절의(예를 들면, 이전 연도의 같은 달) 마지막 관측값으로 둡니다. $T+h$ 에 대한 예측을 식으로 쓰면,

$$\hat{y}_{T+h|T} = y_{T+h-m(k+1)},$$

- m 은 계절성의 주기(seasonal period), $k=\lfloor(h-1)/m\rfloor+1$, $\lfloor u \rfloor$ 은 u 의 정수 부분을 나타냅니다.
 - 예를 들면, 월별 데이터에서 미래의 모든 2월 값들의 예측값이 마지막으로 관측한 2월 값과 같습니다. 분기별 데이터에서, 미래의 모든 2분기 값의 예측치가 마지막으로 관측한 2분기 값과 같습니다.

```
snaive(y, h)
```

3.1 몇 가지 단순한 예측 기법

- 표류 기법
 - 단순 기법(naïve method)을 수정하여 예측값이 시간에 따라 증가하거나 감소하게 할 수 있습니다.
 - 여기에서 (**표류(drift)**)라고 부르는) 시간에 따른 변화량을 과거 데이터에 나타나는 평균 변화량으로 정합니다. 그러면 $T+h$ 시간에 대한 예측값은 다음과 같이 주어집니다.

$$\hat{y}_{T+h|T} = y_T + \frac{h}{T-1} \sum_{t=2}^T (y_t - y_{t-1}) = y_T + h \left(\frac{y_T - y_1}{T-1} \right).$$

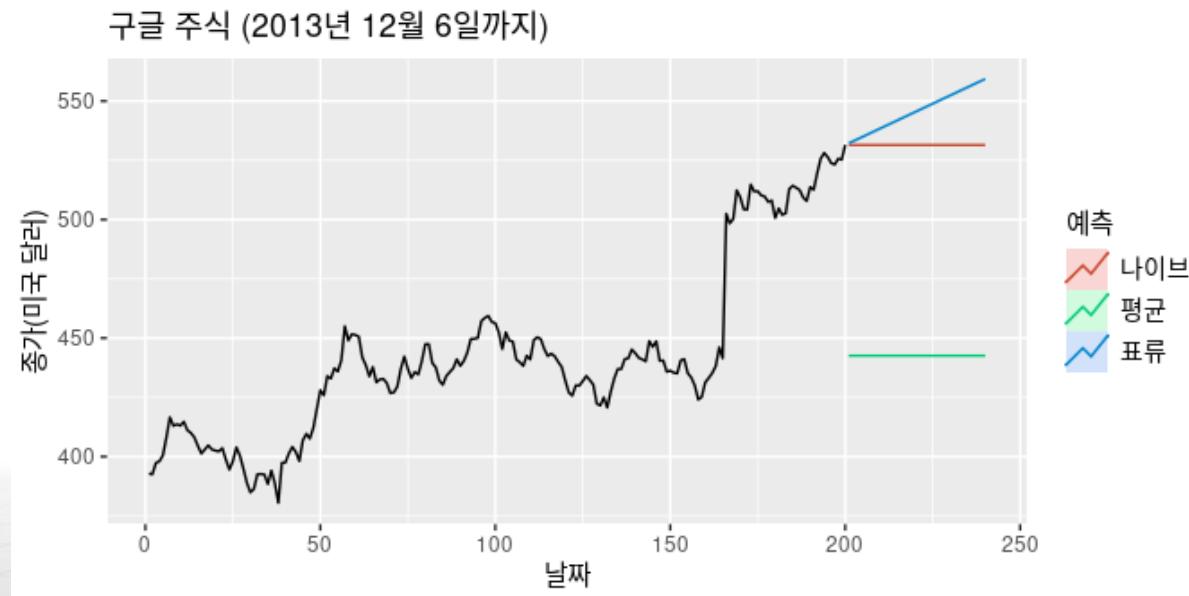
- 처음과 마지막 관측값에 선을 긋고 이 선을 미래로 외삽(extrapolation)한 것과 같습니다.

```
rwf(y, h, drift=TRUE)
```

3.1 몇 가지 단순한 예측 기법

- 예제: 구글 주식 일별 종가

- 때때로 이런 단순한 기법 중의 하나가 사용할 수 있는 가장 좋은 예측 기법이 될 수 있습니다.
- 하지만, 많은 경우에, 이러한 기법은 선택한 기법보다는 벤치마크 역할을 할 것입니다.
 - 즉, 새로운 기법이 이런 단순한 기법보다 좋은지 확인하기 위해, 우리가 다룰 모든 예측 기법을 이런 단순한 기법과 비교할 것입니다. 그렇지 않으면, 그 새로운 기법은 고려하지 않는 것이 낫습니다.



200일 간의 구글 주식 일별 종가에 기초한 예측값

3.3 잔차 진단

- 잔차
 - 시계열 모델에서 “잔차(Residuals)”는 모델을 맞춘 후에 남는 것을 의미합니다. 다양한 (하지만 전부는 아닌) 시계열 모델에서, 잔차(residual)는 관측값과 대응되는 적합값(fitted value)과 관측값의 차이와 같습니다.

$$e_t = y_t - \hat{y}_t.$$

- 잔차(residual)는 어떤 모델이 데이터의 정보를 적절하게 잡아냈는지 여부를 확인할 때 유용합니다. 좋은 예측 기법은 다음과 같은 특징을 갖는 잔차(residual)를 낼 것입니다.
 1. 잔차(residual)에 상관 관계가 없습니다. 잔차 사이에 상관관계(correlation)가 있다면, 잔차에 예측값을 계산할 때 사용해야하는 정보가 남아 있는 것입니다.
 2. 잔차의 평균이 0입니다. 잔차의 평균이 0이 아니라면, 예측값이 편향(bias)될 것입니다.
 3. 잔차의 분산이 상수입니다.
 4. 잔차가 정규 분포를 따릅니다.

3.3 잔차 진단

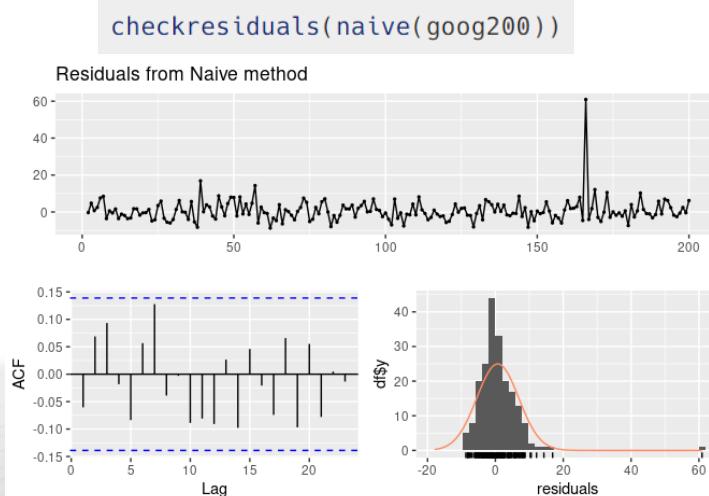
- 자기상관에 대한 포트맨토 검정
 - 박스-피어스(Box-Pierce) 검정

$$Q = T \sum_{k=1}^h r_k^2,$$

h: 고려할 최대 시차(lagged value)
 T: 관측값의 개수
 K: 모델의 매개변수. 모델의 잔차에서 계산하는 대신 원본 데이터에서 이것을 계산한다면 K=0
 Degree of freedom: h-K

- 융-박스(Ljung-Box) 검정

$$Q^* = T(T + 2) \sum_{k=1}^h (T - k)^{-1} r_k^2.$$



```

#>
#> Ljung-Box test
#>
#> data: Residuals from Naive method
#> Q* = 11, df = 10, p-value = 0.4
#>
#> Model df: 0. Total lags used: 10
  
```

3.4 예측 정확도 평가

- 학습 데이터와 테스트 데이터
 - 정답지를 사용하여 예측 정확도를 평가하는 것은 중요합니다. 결과적으로, 잔차(residual)의 크기는 참 예측 오차(forecast error)가 얼마나 클 지에 대해 믿을만한 지표가 아닙니다.
 - 예측치의 정확도는 모델이 모델을 맞출 때 사용하지 않은 새로운 데이터를 얼마나 잘 맞추는지 여부로만 결정할 수 있습니다.
 - 흔히 사용할 수 있는 데이터를 **학습 데이터(training data)**와 **테스트 데이터(test data)** 이렇게 두 부분으로 나눕니다
 - 테스트 데이터(test data)의 크기는 표본이 얼마나 긴지와 얼마나 멀리 예측할 지에 따라 달라지긴 하지만 보통은 전체 표본의 약 20% 정도입니다



3.4 예측 정확도 평가

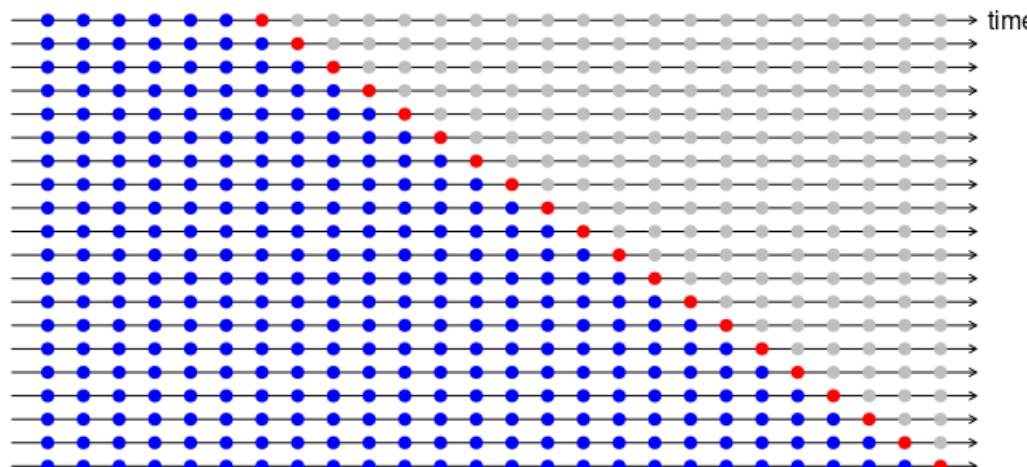
- 눈금에 의존하는 오차
 - 예측 오차(forecast error)는 데이터와 같은 눈금 위에 있습니다. 따라서 e_t 만 고려하는 정확도 측정값은 눈금(scale)에 의존하고 다른 단위(unit)를 포함하는 시계열을 비교하는데 사용할 수 없습니다.

평균 절대 오차(Mean absolute error): $MAE = \text{mean}(|e_t|)$,

제곱근 평균 제곱 오차(Root mean squared error): $RMSE = \sqrt{\text{mean}(e_t^2)}$.

3.4 예측 정확도 평가

- 시계열 교차 검증
 - 시계열 교차 검증(cross-validation)은 데이터를 더 세련되게 학습/테스트하는 방법입니다.
 - 대응되는 학습(training) 데이터는 테스트(test) 데이터를 구성하는 관측에 앞서(prior) 일어난 관측만으로 구성됩니다. 그래서 예측치를 구성할 때 미래 측정치를 전혀 사용하지 않습니다. 작은 학습 데이터에서 신뢰할만한 예측을 얻을 수 없기 때문에, 초반부의 관측값을 테스트 데이터로 고려하지 않습니다.
 - 다음의 도표는 학습 데이터와 테스트 데이터를 연속하여 나타낸 것입니다. 여기에서 파란 관측값은 학습 데이터, 빨간 관측값은 테스트 데이터입니다.



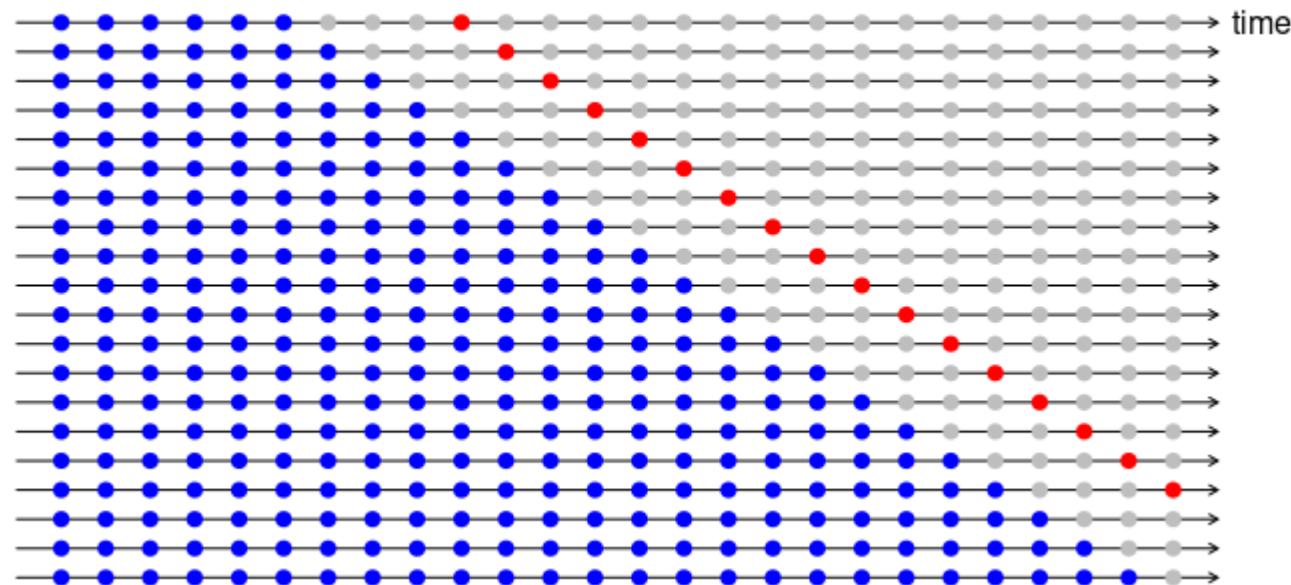
- 예측 정확도(forecast accuracy)는 테스트(test) 데이터에 대한 평균으로 계산합니다. 예측하는 원점(origin)을 시간에 따라 앞으로 굴리기 때문에 때때로 이 과정을 "예측 원점 굴리기에 대한 평가(evaluation on a rolling forecasting origin)"라고도 합니다.

3.4 예측 정확도 평가

• 시계열 교차 검증

- 시계열 예측에서, 한 단계 예측치는 여러 단계 예측치와 그렇게 관련이 있지 않을 수도 있습니다.

- 이런 경우에는, 예측 원점 굴리기(rolling forecasting origin)에 기초한 교차 검증(cross-validation) 과정을 여러 단계 오차(multi-step forecast)를 사용할 수 있도록 변형할 수 있습니다.



3.6 R의 forecast 패키지

- R의 forecast 패키지
 - forecast 객체를 결과로 내는 함수들:
 - meanf()
 - naive(), snaive()
 - rwf()
 - croston()
 - stlf()
 - ses()
 - holt(), hw()
 - splinef()
 - thetaf()
 - forecast()

시계열 데이터를 이용한 회귀 분석 이해

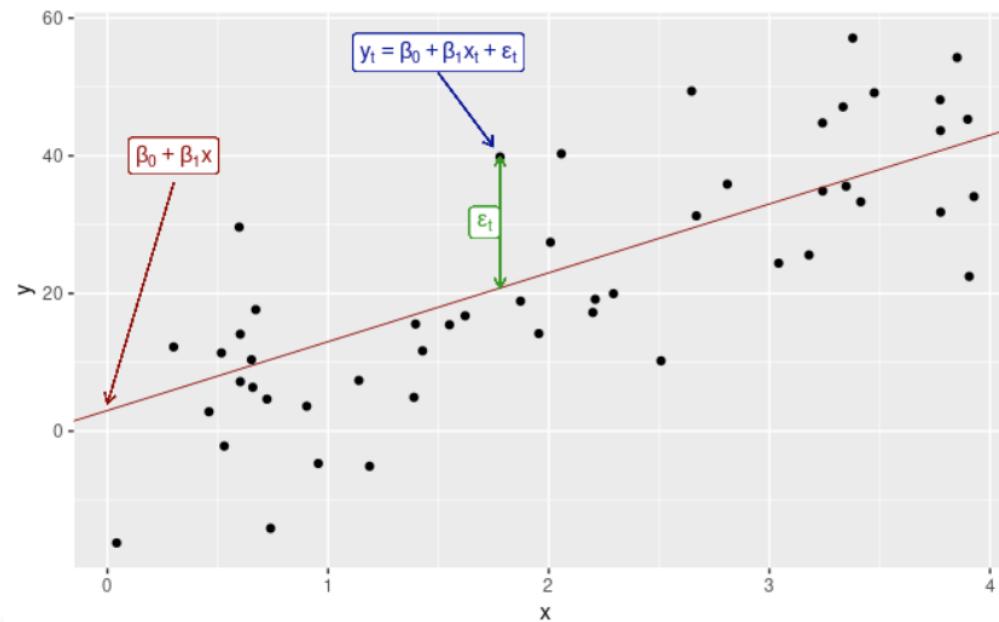
- 시계열 회귀 모델

- 이 장에서는 회귀 모델(regression model)을 다룹니다. 시계열 y 을 예측할 때 이것이 다른 시계열 x 와 선형 관계가 있다고 가정하는 것이 기본 개념입니다.
 - 예를 들면, 전체 광고 비용 x 을 예측변수(predictor variable)로 두고 월별 판매량 y 을 예측하는 상황이 있을 수 있습니다. 또는 온도 x_1 과 요일 x_2 을 예측변수로 사용하여 일별 전력 수요 y 를 예측하는 상황이 있을 수 있습니다.
- **목표 예상변수(forecast variable)** y 는 때때로 회귀선, 종속 변수, 또는 피설명 변수라고 부르기도 합니다.
- **예측변수(predictor variables)** x 는 종종 회귀자, 독립 변수, 설명 변수라고 부르기도 합니다. 이 책에서는 항상 y 를 “목표 예상 (forecast)” 변수로 x 를 “예측(predictor)” 변수로 부를 것입니다.

4.1 선형 모델

- 단순 선형 회귀
 - 가장 간단한 경우에는, 회귀 모델이 목표 예상변수(forecast variable) y 와 하나의 예측변수(predictor variable) x 사이의 선형 관계를 다룹니다.

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t.$$



4.1 선형 모델

- 다중 선형 회귀
 - 두 개 이상의 예측변수(predictor variable)가 있을 때, 모델을 **다중 회귀 모델(multiple linear regression)**이라고 부릅니다. 다중 회귀 모델의 일반적인 형태는 다음과 같습니다.

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_k x_{k,t} + \varepsilon_t,$$

- 여기에서 y 는 예측될 목표 예상변수(forecast variable)이고, x_1, \dots, x_k 들은 k 개의 예측변수(predictor variable)입니다.
- 예측변수(predictor variable) 각각은 숫자 형태이어야만 합니다.
- 계수 β_1, \dots, β_k 은 모델에서 다른 모든 예측변수(predictor variable)의 효과를 고려한 후의 각 예측변수의 효과를 나타냅니다. 따라서, 계수는 예측변수(predictor variable)의 **한계 효과**(marginal effects)를 나타냅니다.

4.1 선형 모델

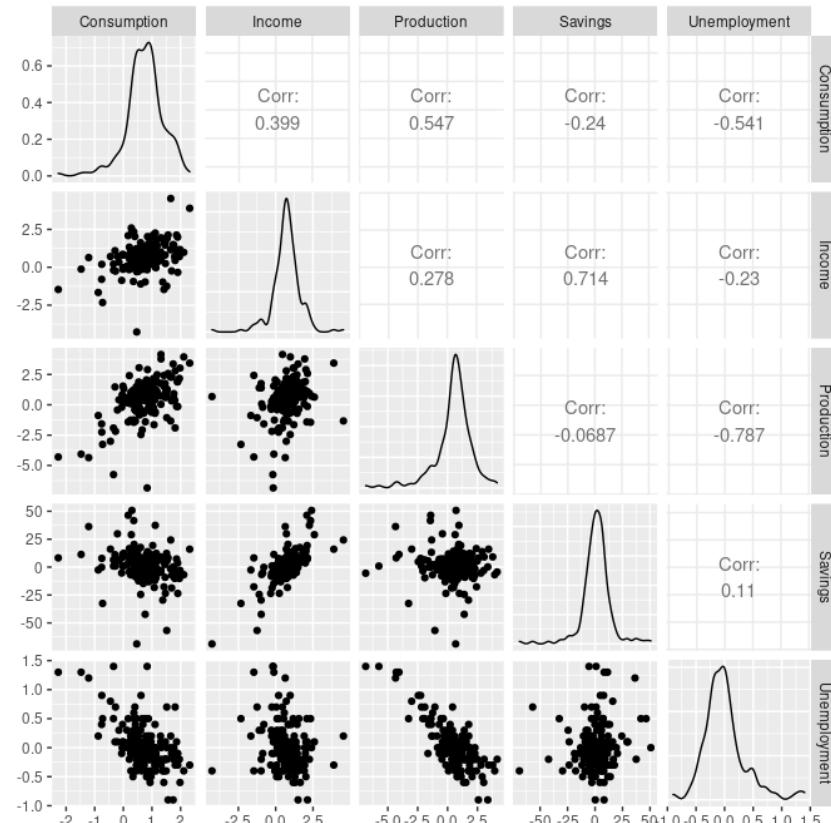
• 다중 선형 회귀

- 미국 산업 생산, 개인 저축, 실업률 분기별 백분율 변화를 1960년 1분기부터 2016년 3분기까지 나타낸 것.



4.1 선형 모델

- 다중 선형 회귀
 - 5개 변수의 산점도 행렬 (scatterplot matrix)
 - 첫 번째 열은 목표 예상변수 - forecast variable - (소비; consumption)와 각 예측변수 (predictor variable) 사이의 관계를 나타냅니다.
 - 산점도 (scatter plot)는 소득과 산업 생산과의 양의 관계, 저축과 실업률과의 음의 관계를 나타냅니다. 이러한 관계의 강도는 첫 번째 행에 따라 상관 계수 (correlation coefficient)로 나타내었습니다.
 - 나머지 산점도와 상관 계수 (correlation coefficient)는 예측변수 (predictor variable) 사이의 관계를 나타냅니다.



- 선형 회귀 가정
 - 선형 회귀 모델을 사용할 때, 다중회귀 수식에서는 변수에 대한 몇 가지 가정을 암시적으로 사용하고 있습니다.
 - 첫 번째, 모델은 현실에 대한 타당한 근사식이고; 즉, 목표 예상변수(forecast variable)와 예측변수(predictor variable) 사이의 관계는 이러한 선형 관계식을 만족한다는 것입니다.
 - 두 번째, 오차(error) ($\varepsilon_1, \dots, \varepsilon_T$)에 대해 다음과 같은 가정을 사용합니다:
 - 오차 평균은 0입니다; 그렇지 않으면, 예측값이 체계적으로 편향될 것입니다.
 - 오차에는 자기상관관계(autocorrelation)가 없습니다; 그렇지 않으면, 데이터에 악용될 수 있는 정보가 더 많아서 예측값이 비효율적으로 될 것입니다.
 - 오차는 예측변수(predictor variable)와 상관관계(correlation)가 없습니다; 그렇지 않은 경우에는 모델의 체계적인 부분에 포함되어야 할 정보가 더 있을 수 있습니다.
 - 예측 구간(prediction interval)을 쉽게 내기 위해 오차가 일정한 분산값 σ^2 을 가지는 정규 분포(normal distribution)를 나타낸다고 가정하는 것도 유용합니다.

통제하는 것은 불가능하고 단순히 관측값 분집니다. 따라서 이러한 가정을 사용합니다.

- 선형 회귀 가정
 - 선형 회귀 모델을 사용할 때, 다중회귀 수식에서는 변수에 대한 몇 가지 가정을 암시적으로 사용하고 있습니다.
 - 첫 번째, 모델은 현실에 대한 타당한 근사식이고; 즉, 목표 예상변수(forecast variable)와 예측변수(predictor variable) 사이의 관계는 이러한 선형 관계식을 만족한다는 것입니다.
 - 두 번째, 오차(error) ($\varepsilon_1, \dots, \varepsilon_T$)에 대해 다음과 같은 가정을 사용합니다:
 - 오차 평균은 0입니다; 그렇지 않으면, 예측값이 체계적으로 편향될 것입니다.
 - 오차에는 자기상관관계(autocorrelation)가 없습니다; 그렇지 않으면, 데이터에 악용될 수 있는 정보가 더 많아서 예측값이 비효율적으로 될 것입니다.
 - 오차는 예측변수(predictor variable)와 상관관계(correlation)가 없습니다; 그렇지 않은 경우에는 모델의 체계적인 부분에 포함되어야 할 정보가 더 있을 수 있습니다.
 - 예측 구간(prediction interval)을 쉽게 내기 위해 오차가 일정한 분산값 σ^2 을 가지는 정규 분포(normal distribution)를 나타낸다고 가정하는 것도 유용합니다.
 - 선형 회귀 모델에서 또 하나의 중요한 가정은 각 예측변수(predictor variable) x 는 확률 변수(random variable)가 아니라는 것입니다. 실험실에서 통제된 실험을 하고 있는 상황에서는, 각 x 값을 통제할 수 있고 (그래서 무작위적일 수 없습니다) 결과 y 값을 관측할 수 있습니다.
 - (비즈니스 및 경제 분야의 대부분의 데이터를 포함한) 관측 데이터로 x 값을 통제하는 것은 불가능하고 단순히 관측할 뿐입니다. 따라서 이러한 가정을 사용합니다.

4.2 최소 제곱 추정

- 최소 제곱 추정
 - 물론 실제로는 관측값을 가지고 있지만 계수 값 β_0, \dots, β_k 은 모릅니다. 이러한 값은 데이터로부터 추정해야 합니다.
 - 최소 제곱 원리(least square principle)는 제곱 오차의 합을 최소화하여 계수를 효과적으로 선택할 수 있는 방법입니다. 즉, 아래 값을 최소화하는 β_0, \dots, β_k 값을 선택합니다.

$$\sum_{t=1}^T \varepsilon_t^2 = \sum_{t=1}^T (y_t - \beta_0 - \beta_1 x_{1,t} - \beta_2 x_{2,t} - \dots - \beta_k x_{k,t})^2.$$

- 오차를 제곱하여 더한 양의 최소값을 나타내기 때문에 이것을 “최소 제곱(least square)” 추정이라고 부릅니다. 계수의 가장 좋은 추정치를 찾는 것을, 종종 모델을 데이터에 “맞춘다”고 부르거나, 때때로 모델을 “학습(learning)”시키거나 “훈련(training)”시킨다고 부릅니다.
- 추정된 계수를 언급할 때, 이러한 $\hat{\beta}_0, \dots, \hat{\beta}_k$ 기호를 사용하여 나타낼 것입니다.

4.2 최소 제곱 추정

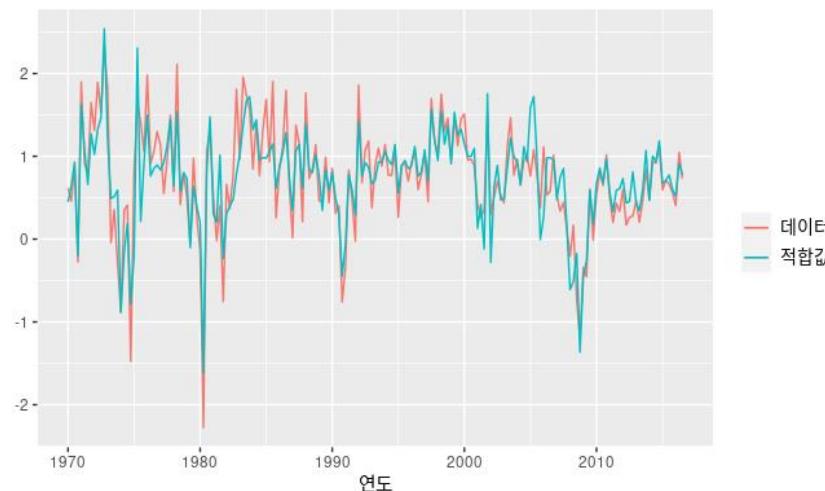
- 적합값 (fitted value)

- 회귀식에서 추정된 계수를 사용하고 오차항을 0으로 두어 y 의 예측값을 얻을 수 있습니다. 일반적으로 다음과 같이 적습니다.

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1,t} + \hat{\beta}_2 x_{2,t} + \cdots + \hat{\beta}_k x_{k,t}.$$

- $x_{1,t}, \dots, x_{k,t}$ for $t = 1, \dots, T$ 값을 대입하면 학습-표본 안의 y_t 예측값을 얻습니다. 이 예측값을 적합값(fitted value)이라고 부릅니다.
 - y 의 참 미래 예측값이 아니고 모델을 추정하기 위해 사용한 데이터의 예측값이라는 것에 주목하시길 바랍니다.

미국 소비 지출의 백분율 변화



4.2 최소 제곱 추정

- 적합도(Goodness-of-fit)
 - 선형 회귀 모델이 데이터에 얼마나 잘 맞는지 요약하는 일반적인 방법은 결정 계수(coefficient of determination) 또는 R^2 를 사용하는 것입니다. 이것은 관측한 y 값과 예측한 \hat{y} 값 사이의 상관관계(correlation)의 제곱으로 계산할 수 있습니다. 대신에, 다음과 같이 계산할 수도 있습니다.
 - 여기에서 합산 기호는 모든 관측값에 대한 것입니다. 따라서, 회귀 모델로 설명되는 목표 예상변수(forecast variable)의 변동 비율을 반영합니다.
 - 예측값이 실제 값과 가깝다면, R^2 는 1에 가까울 것입니다. 반면에, 예측값이 실제 값과 관련이 없다면(다시 말해, 절편이 있다고 가정하면), $R^2 = 0$ 입니다. 모든 경우에 R^2 는 0과 1 사이의 값을 갖습니다.

$$R^2 = \frac{\sum(\hat{y}_t - \bar{y})^2}{\sum(y_t - \bar{y})^2}, \quad \bar{y} = \frac{\sum y}{n}$$

4.2 최소 제곱 추정

- 회귀 분석의 표준 오차

- 모델이 데이터에 얼마나 잘 들어 맞았는지에 대한 척도는 “잔차 표준 오차(residual standard error)”라고 하는 잔차(residual)의 표준 편차입니다.

$$\hat{\sigma}_e = \sqrt{\frac{1}{T - k - 1} \sum_{t=1}^T e_t^2}.$$

- 여기에서 k 는 모델에 포함된 예측변수(predictor variable)의 수입니다. 잔차(residual)를 계산할 때 $k + 1$ 개의 매개변수(각 예측 변수에 대한 절편과 계수)를 추정했기 때문에 $T - k - 1$ 로 나누는 것에 주목하시길 바랍니다.
- 표준 오차는 모델이 내는 평균 오차의 크기와 관련이 있습니다. 이 오차를 y 의 표본 평균이나 모델의 정확도에 대해 몇 가지 파악하기 위해 y 의 표준 편차와 비교할 수 있습니다.

4.3 회귀 모델 평가

- 회귀 모델 평가
 - 관측된 y 값과 해당하는 적합값 \hat{y} 사이의 차이값은 $t = 1, \dots, T$ 에 대해 다음과 같이 학습-모음 오차 또는 “잔차(residual)”로 정의합니다.

$$\begin{aligned} e_t &= y_t - \hat{y}_t \\ &= y_t - \hat{\beta}_0 - \hat{\beta}_1 x_{1,t} - \hat{\beta}_2 x_{2,t} - \cdots - \hat{\beta}_k x_{k,t} \end{aligned}$$

- 각 잔차(residual)는 관련된 관측값의 예측할 수 없는 성분입니다.
- 잔차(residual)에는 다음 두 가지를 포함하여 몇 가지 유용한 성질이 있습니다:

$$\sum_{t=1}^T e_t = 0 \quad \text{and} \quad \sum_{t=1}^T x_{k,t} e_t = 0 \quad \text{for all } k.$$

- 이러한 성질의 결과로, 잔차(residual)의 평균이 0이고 잔차(residual)와 예측변수(predictor variable)에 대한 관측값 사이의 상관관계(correlation)도 0이라는 것이 분명합니다. (Y절편이 모델에서 생략될 때, 이것이 반드시 참은 아닙니다.)

4.3 회귀 모델 평가

- 잔차의 ACF 그래프
 - 시계열 데이터에서, 현재 시점에 관측한 변수의 값은 이전 기간의 값과 비슷하거나 그보다 이전 기간의 값과 비슷할 것입니다.
 - 따라서 회귀 모델로 시계열 데이터를 맞출 때, 보통은 잔차의 자기상관(autocorrelation)을 찾는 작업을 합니다
 - 이 경우에, 추정된 모델 오차에서 자기상관관계(autocorrelation)가 없다는 가정을 위배하여 예측이 비효율적일 수 있습니다. 이것은 더 나은 예측을 위해 모델에서 고려해야 할 정보가 더 있다는 것을 의미합니다.
 - 자기상관(autocorrelation)을 가지는 오차가 있는 모델로 낸 예측치는 여전히 편향되어있지는 않아서 “잘못된” 것은 아니지만, 보통 필요 한 것보다 더 큰 예측구간(prediction interval)을 가질 것입니다.
 - 따라서 항상 잔차의 ACF (AutoCorrelation Function) 그래프를 살펴봐야 합니다.

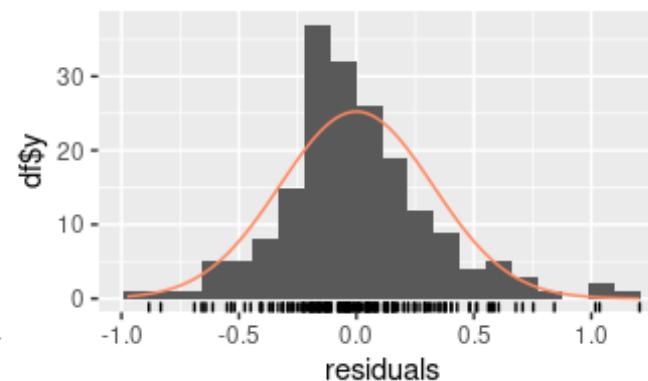
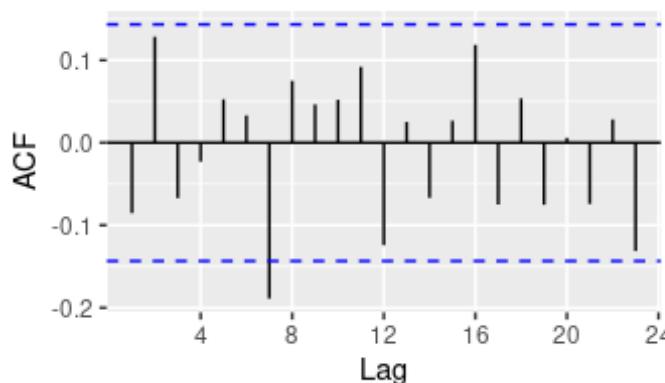
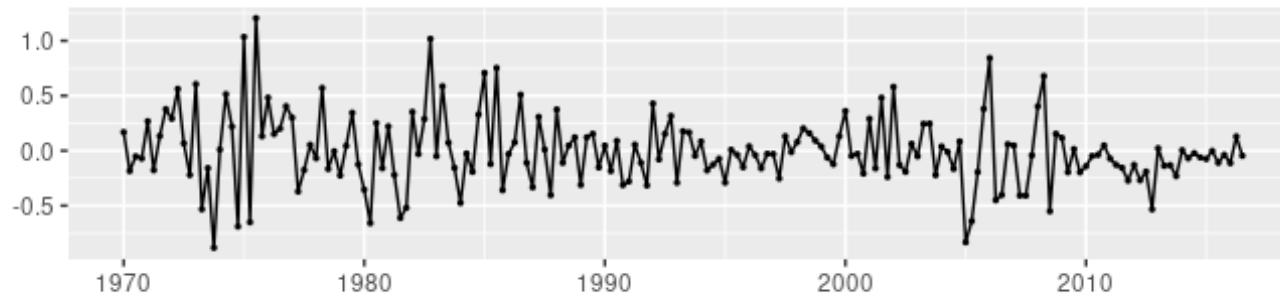
4.3 회귀 모델 평가

- 잔차의 ACF 그래프
 - 회귀 모델을 고려하여 설계한 잔차(residual)의 자기상관(autocorrelation)을 확인하는 유용한 또 다른 방법은 브로이쉬-갓프레이(Breusch-Godfrey) 검정입니다.
 - 이 방법은 계열 상관(serial correlation)에 대한 LM (Lagrange Multiplier) 검정이라고도 합니다.
 - 이것은 어떤 특정한 순서까지 잔차에 자기상관(autocorrelation)이 없다는 결합 가설(joint hypothesis)을 검증하는데 사용합니다.
 - 작은 p-값은 잔차(residual)에 중요한 자기상관(autocorrelation)이 남아 있다는 것을 나타냅니다.
 - 브로이쉬-갓프레이(Breusch-Godfrey) 검정은 융-박스(Ljung-Box) 검정과 비슷합니다만, 회귀 모델을 위해 설계되었다는 점이 구체적인 차이입니다.
- 잔차의 히스토그램
 - 잔차가 정규 분포를 따르는지 항상 확인하는 것은 좋은 접근 방식입니다. 이전에 설명한 것처럼, 이러한 작업은 예측 작업에서 필수적인 것은 아닙니다만, 이를 통해 예측 구간(prediction interval)을 훨씬 쉽게 계산할 수 있게 됩니다.

4.3 회귀 모델 평가

- 미국 분기별 소비에 대한 회귀 모델로부터 얻은 잔차를 분석하기

Residuals from Linear regression model



```
#> Breusch-Godfrey test for serial correlation of  
#> order up to 8  
#>  
#> data: Residuals from Linear regression model  
#> LM test = 15, df = 8, p-value = 0.06
```

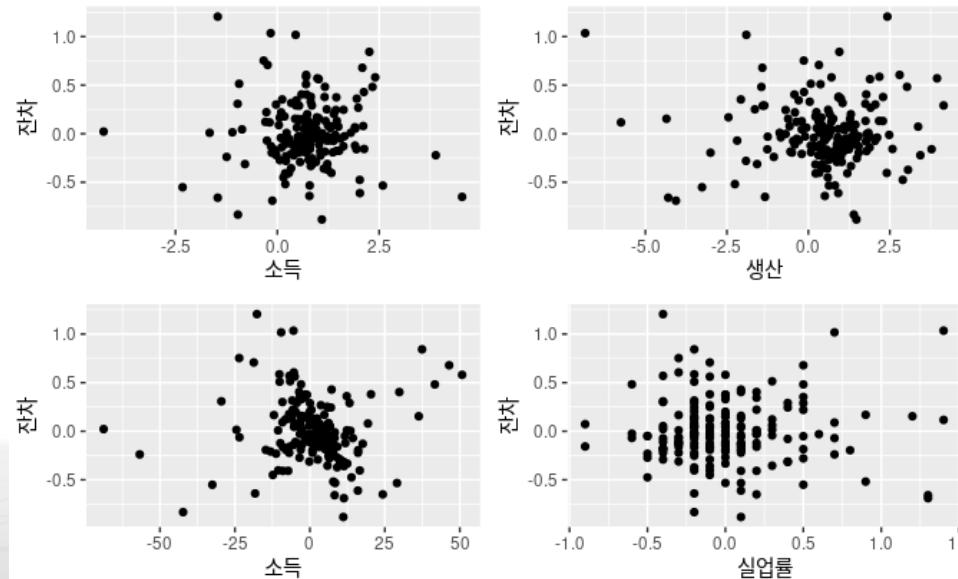
- 시간 그래프(time plot)에서 몇몇 부분은 시간에 따라 변하지만 나머지는 상대적으로 두드러지진 않습니다. 이러한 이분산성(heteroskedasticity)이 예측 구간 범위를 정확하지 않게 만들 수도 있습니다

- 히스토그램은 잔차가 살짝 치우친 것 같다는 것을 보여 줍니다. 이것은 예측 구간의 범위 확률에 영향을 줄 수도 있습니다.

4.3 회귀 모델 평가

• 예측변수에 대한 잔차 그래프

- 잔차가 체계적인 패턴을 보이지 않고 무작위로 흩뿌려진 형태로 나타날 것 같습니다.
- 이것을 확인하는 단순하고 빠른 방식은 각 예측변수(predictor variable)에 대한 잔차(residual)의 산점도(scatterplot)를 살펴보는 것입니다.
- 이러한 산점도(scatterplot)에 패턴이 나타나면, 관계가 비선형적일 수 있어서 그에 따라 모델을 수정해야합니다.

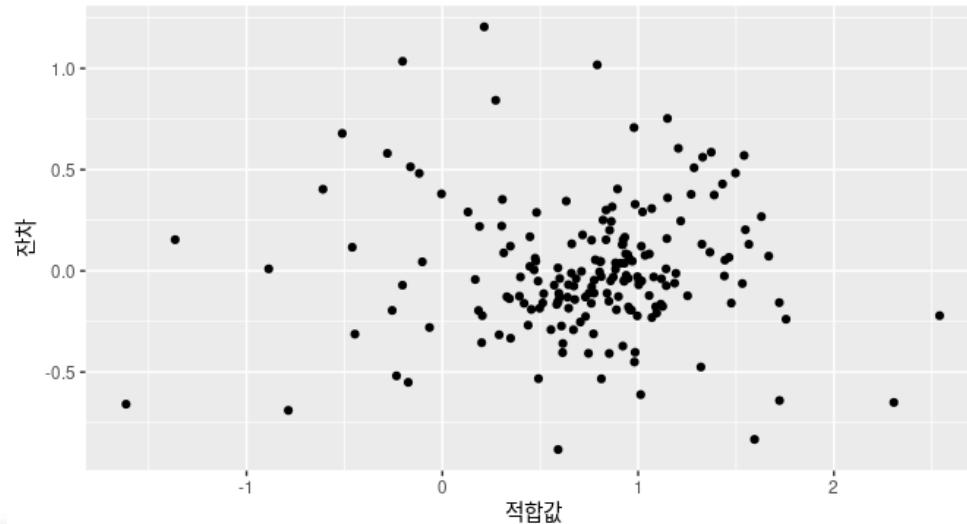


- 각 예측변수(predictor variable)에 대해 나타낸, 미국 소비 예측에 대한 다중 회귀 모델에서 얻은 잔차는 무작위적으로 흩뿌려진 것처럼 보입니다.
- 따라서 이 경우에 만족 할 만한 결과를 얻었습니다.

4.3 회귀 모델 평가

• 적합값에 대한 잔차 그래프

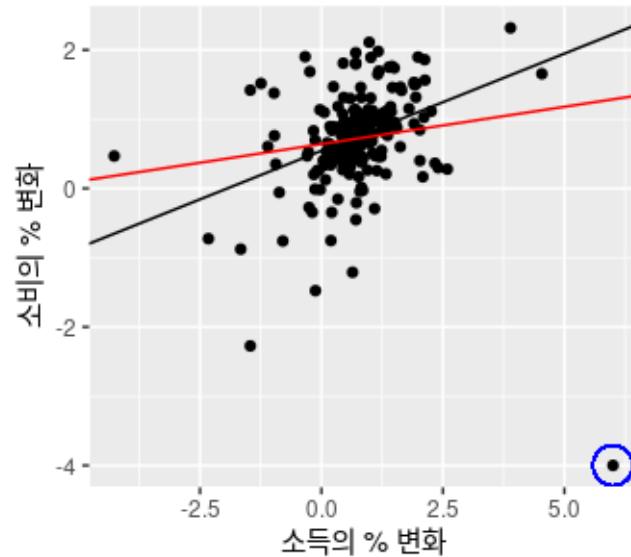
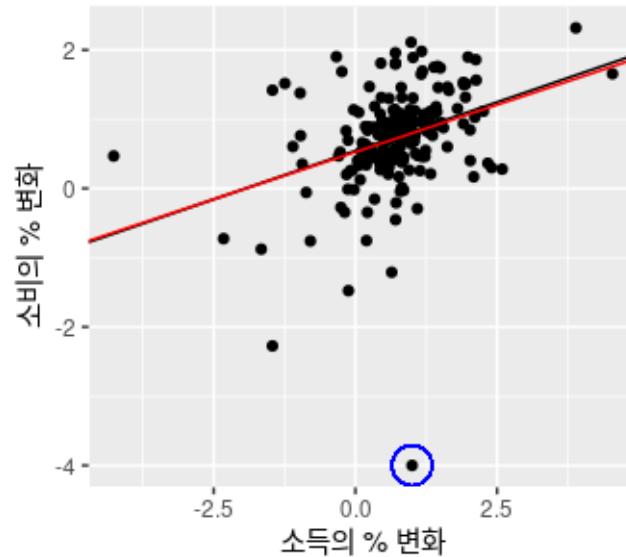
- 적합값에 대해 잔차를 그려도 어떠한 패턴도 나타나지 않아야 합니다. 패턴을 보았다면, 오차에 “이분산성(heteroscedasticity)”이 있을 수 있습니다.
- 이는 잔차의 분산이 일정하지 않을 수도 있다는 것을 의미합니다. 이러한 문제가 생기면, 목표 예상변수(forecast variable)에 로그나 제곱근 같은 변환을 취해야 할 수도 있습니다.



- 이전 예제의 연장선상에서, 왼쪽그림은 잔차를 적합값에 대해 나타냅니다.
- 무작위 산점도는 오차에 등분산성(homoscedasticity)이 있다는 것을 말해줍니다.

4.3 회귀 모델 평가

- 이상값과 영향력 있는 관측값
 - 대다수의 데이터에 비해 극단적인 값을 갖는 관측값을 **이상값 (outliers)**이라고 부릅니다. 회귀 모형의 추정된 계수에 큰 영향을 주는 관측값을 **영향력 있는 관측값**이라고 부릅니다.
 - 보통, 영향력 있는 관측값은 x 방향에서 극단적인 값을 갖는 이상값이기도 합니다.

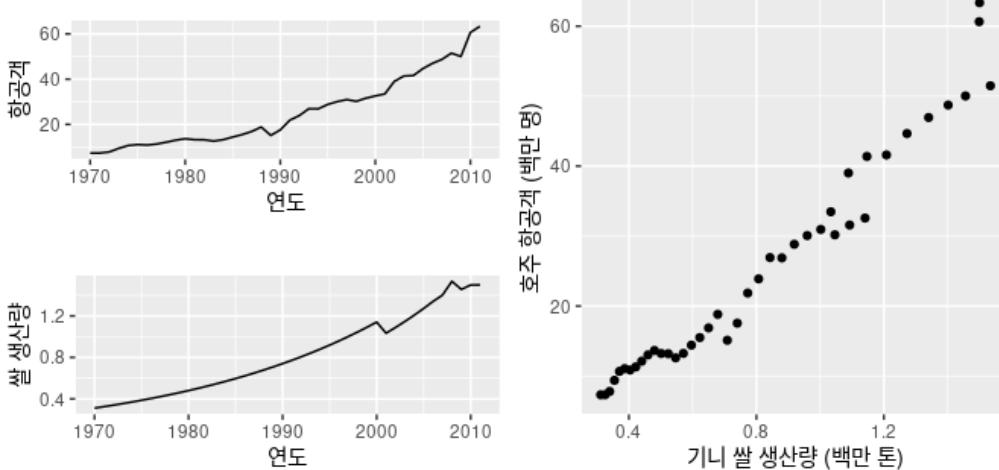


이상값을 제거한 경우와 그렇지 않은 경우 모두 결과를 보고하는 것이 현명합니다.

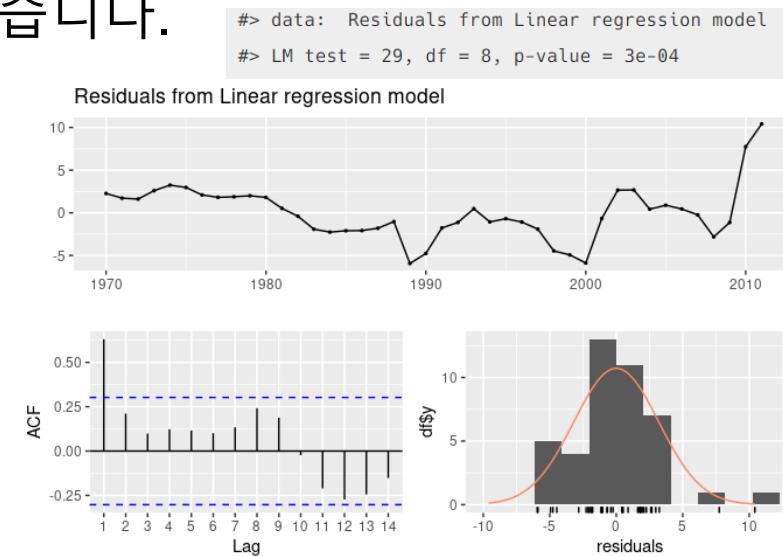
4.3 회귀 모델 평가

• 허위회귀(Spurious regression)

- 대부분의 경우에는, 시계열 데이터에 “정상성(stationarity)이 나타나지 않습니다”; 즉, 시계열의 값이 일정한 평균이나 일정한 분산으로 변하지 않습니다.
- 정상 시계열(stationary time series)을 다룰 것입니다만, 여기에서는 정상성(stationarity)이 나타나지 않는 데이터가 회귀 모델에 어떤 영향을 미칠 수 있는지 설명하겠습니다.



둘 다 같은 방식으로 상향 추세이기 때문에 단순하게 관련 있는 것처럼 보입니다. 하지만, 호주의 승객 수송량은 기니의 쌀 생산량과 관련이 없습니다.



높은 R^2 와 높은 잔차(residual) 자기 상관은 허위회귀(spurious regression)라는 것을 나타내는 신호가 될 수 있음

4.4 예측변수 선택

- 예측변수 선택
 - 가능한 예측변수(predictor variable)가 많이 있을 때는, 회귀 모델에서 사용할 가장 좋은 예측변수(predictor variable)를 선택하는 몇 가지 전략이 필요합니다.
 - 추천하지 않는 일반적인 접근 방식은 목표 예상변수(forecast variable)를 특정한 예측변수(predictor variable)에 대해 그래프로 나타내서 눈에 띄는 관계가 없는 경우에 버리는 것입니다.
 - 또한 유효하지 않은 또 다른 일반적인 접근 방식은 모든 예측변수(predictor variable)에 대해 다중 선형 회귀 분석을 하여 p -값들이 0.05보다 큰 모든 변수를 무시하는 것입니다.
 - 예측하는 것이 목표가 아닐 때도, 2개 이상의 예측변수(predictor variable)가 서로 관련되어있을 때 p -값들이 오해를 일으킬 수 있기 때문에 이것은 좋은 전략이 아닙니다

- 조절된 R^2
 - 컴퓨터에 나오는 회귀 분석 출력에는 항상 다른 R^2 값이 나옵니다.
 - 하지만, 이 값은 모델의 예측 능력을 측정하는 좋은 방법이 아닙니다. 이를 통해 모델이 과거 데이터를 얼마나 잘 설명하는지 측정할 수 있지만, 모델이 미래 데이터를 얼마나 잘 예측할지를 측정할 수는 없습니다.
 - R^2 는 “자유도(degree of freedom)”를 허용하지 않습니다. 어떤 변수를 추가하면 변수가 적합하지 않더라도 R^2 의 값이 증가하는 경향이 있습니다. 이러한 이유로 예측하는 사람들은 R^2 를 모델이 좋은 예측값을 낼지 판단하기 위해 사용하지 않아야 합니다. 과대적합(over-fitting)이 발생할 수 있기 때문입니다.
 - 제곱 오차의 최소 합(SSE)을 내는 모델을 고르는 것도 위와 같은 개념입니다. 여기에서 SSE는 다음과 같은 양입니다.

$$SSE = \sum_{t=1}^T e_t^2.$$

- SSE를 최소화하는 것은 R^2 를 최대화하는 것과 같고, 항상 변수가 가장 많은 모델을 선택할 것이므로 예측변수(predictor variable)를 선택하는 올바른 방법이 아닙니다.

4.4 예측변수 선택

- 조절된 R^2

- 이러한 문제를 해결하기 위해 설계된 방법은 (“R-바-제곱(R-bar-squared)”이라고도 부르는) 조정된 R^2 입니다.

$$\bar{R}^2 = 1 - (1 - R^2) \frac{T - 1}{T - k - 1},$$

T 는 관측값의 수이고 k 는 예측변수(predictor variable)의 수

- 각각의 추가된 예측변수(predictor variable)에 따라 증가하지 않을 것이기 때문에 R^2 가 개선된 것입니다. 이 측정량을 사용하면 \bar{R}^2 가 가장 큰 값을 갖는 것이 가장 좋은 모델이 될 것입니다.
 - \bar{R}^2 최대화하는 것은 주어지는 표준 오차 $\hat{\sigma}_e$ 를 최소화하는 것과 같습니다.

$$\hat{\sigma}_e = \sqrt{\frac{1}{T - k - 1} \sum_{t=1}^T e_t^2}.$$

4.4 예측변수 선택

- 예측 정확도 측정량을 소개
 - 교차검증 (cross-validation)
 - 아카이케의 정보기준 (Akaike's Information Criterion; AIC)
 - 수정된 아카이케의 정보기준 (corrected AIC; AIC_c)
 - 슈바르츠의 베이지안 정보기준 (SBIC, BIC, 또는 SC)

4.4 예측변수 선택

- 교차검증 (cross-validation)
 - 회귀 모델에서 예측변수(predictor variable)를 선택하기 위해 고전적인 단일 관측치 제거법(leave-one-out) 교차검증을 사용하는 것도 가능합니다. 이 방법은 더 빠르면서 데이터를 더 효과적으로 사용할 수 있도록 합니다. 사용 과정은 다음과 같습니다:
 1. 관측값 t 를 데이터 모음에서 제거하고, 모델로 나머지 데이터를 맞춥니다. 그리고 생략한 관측값에 대한 오차($e_t^* = y_t - \hat{y}_t$)를 계산합니다. (\hat{y}_t 값을 추정할 때 t 번째 관측값이 사용되지 않았기 때문에 잔차와 같지 않습니다.)
 2. $t = 1, \dots, T$ 에 대해 1번 과정을 반복합니다.
 3. e_1^*, \dots, e_T^* 로부터 MSE를 계산합니다. 이것을 CV라고 부릅니다.

4.4 예측변수 선택

- 아카이케의 정보 기준 (AIC)

- 밀접하게 관련된 기법은 아카이케의 정보 기준입니다(Akaike's Information Criterion; AIC). 이 양의 정의는 다음과 같습니다.

$$AIC = T \log\left(\frac{SSE}{T}\right) + 2(k + 2),$$

- T 는 추정에 사용하는 관측값의 수이고 k 는 모델에 있는 예측변수 (predictor variable)의 수입니다.
 - 서로 다른 컴퓨터 패키지 모두 다 선택하게 될 같은 모델을 내야 하긴 하지만, 패키지마다 AIC를 계산할 때 살짝 다른 정의를 사용합니다. 식의 $k + 2$ 부분은 모델에 $k + 2$ 개의 매개변수가 있기 때문에 나타나는 것입니다.
 - 예측변수(predictor variable)에 대한 k 개의 계수, 절편, 잔차에 대한 분산 때문에 $k + 2$ 개입니다.
 - 추정할 필요가 있는 매개변수의 수로 모델을 맞추는 작업(SSE)에 제한을 거는 것이 여기에서 기본적인 개념입니다.
 - AIC 값이 최소인 모델은 예측할 때 종종 가장 좋은 모델이 됩니다. 더 큰 T 값들의 경우에, AIC를 최소화하는 것은 CV 값을 최소화하는 것과 같습니다.

4.4 예측변수 선택

- 수정된 아카이케의 정보 기준 (AIC_c)
 - 작은 T 값들의 경우에, AIC는 예측변수(predictor variable)를 너무 많이 고르는 경향이 있어서 다음과 같이 편향을 수정한 AIC를 사용합니다.

$$AIC_c = \text{AIC} + \frac{2(k+2)(k+3)}{T-k-3}.$$

- AIC처럼, AIC_c 도 최소화해야 합니다.

4.4 예측변수 선택

- 슈바르츠의 베이지안 정보 기준 (SBIC, BIC, 또는 SC)
 - 관련된 측정량은 슈바르츠(Schwarz)의 베이지안(Bayesian) 정보 기준입니다
 - AIC처럼, 가장 좋은 모델을 얻기 위해 BIC를 최소화해야합니다. BIC로 고른 모델은 AIC로 고른 것이나 더 적은 수의 항을 고려한 것과 같습니다. BIC가 AIC보다 매개변수의 수에 더 큰 제한을 주기 때문입니다.
 - 큰 T 값에 대해, BIC를 최소화하는 것은 $v = T[1 - \frac{1}{\log(T)-1}]$ 때 v 개 관측치 제거법 교차 검증하는 것과 비슷합니다.

4.4 예측변수 선택

- 어떤 양을 사용해야 합니까?
 - \bar{R}^2 이 널리 사용되고, 다른 양보다 더 오래되었지만, 이 양은 너무 많은 예측변수(predictor variable)를 고르는 경향이 있어서 예측 작업에 덜 적합합니다.
 - 많은 통계학자는 진짜 근본 모델이 있을 때, 충분한 데이터가 있다면 BIC가 그 근본 모델을 고를 것이기 때문에 BIC를 사용합니다.
 - 하지만, 실제로는 진정한 근본 모델이 있는 경우는 거의 없으며 진정한 근본 모델이 있더라도 (매개변수 추정값이 정확하지 않을 수 있기 때문에) 해당 모델을 선택하는 것이 반드시 가장 좋은 예측치를 낼 것이라고 보장하지 않습니다.
 - 결과적으로, 여기에서는 예측을 목적으로 하는 (큰 T 값에 대해 동등한 모델을 내는) AIC_c , AIC, 또는 CV 통계를 주로 사용하겠습니다.

- 예제: 미국 소비

– 미국 소비 예측 다중 회귀 예제에서 4개의 예측변수(predictor variable)를 고려했습니다.

- 4개의 예측변수(predictor variable)를 고려하면, $2^4=16$ 개의 모델이 가능합니다.
- 이제 4개의 모든 예측변수(predictor variable)가 실제로 유용한지, 이 중에서 하나 이상을 버릴지 여부를 확인할 수 있습니다

소득	생산	저축	실업률	CV	AIC	AICc	BIC	AdjR2
1	1	1	1	0.116	-409.3	-408.8	-389.9	0.749
1	0	1	1	0.116	-408.1	-407.8	-391.9	0.746
1	1	1	0	0.118	-407.5	-407.1	-391.3	0.745
1	0	1	0	0.129	-388.7	-388.5	-375.8	0.716
1	1	0	1	0.278	-243.2	-242.8	-227.0	0.386
1	0	0	1	0.283	-237.9	-237.7	-225.0	0.365
1	1	0	0	0.289	-236.1	-235.9	-223.2	0.359
0	1	1	1	0.293	-234.4	-234.0	-218.2	0.356
0	1	1	0	0.300	-228.9	-228.7	-216.0	0.334
0	1	0	1	0.303	-226.3	-226.1	-213.4	0.324
0	0	1	1	0.306	-224.6	-224.4	-211.7	0.318
0	1	0	0	0.314	-219.6	-219.5	-209.9	0.296
0	0	0	1	0.314	-217.7	-217.5	-208.0	0.288
1	0	0	0	0.372	-185.4	-185.3	-175.7	0.154
0	0	1	0	0.414	-164.1	-164.0	-154.4	0.052
0	0	0	0	0.432	-155.1	-155.0	-148.6	0.000

- “1”은 모델에 포함된 예측변수
- “0”은 모델에 포함되지 않은 예측변수

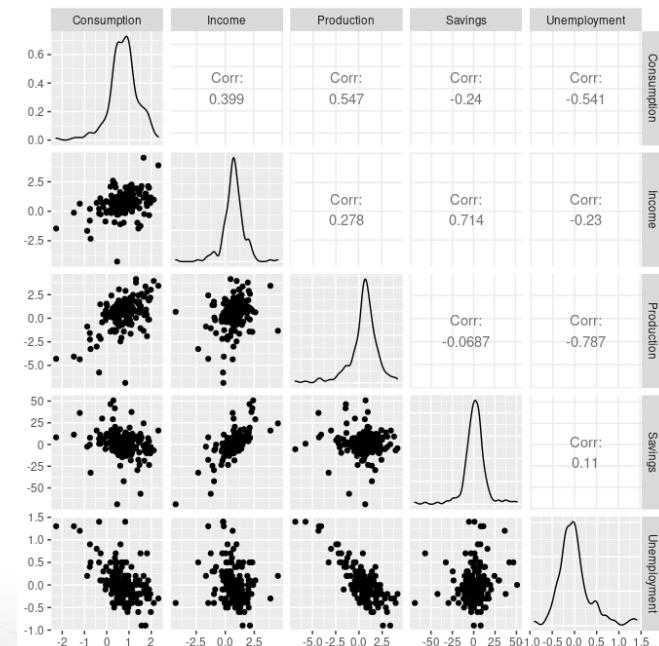
4.4

예측변수 선택

• 예제: 미국 소비

- 소득과 저축이 생산과 실업률보다 더 중요한 변수라는 것을 나타냅니다. 또한, 처음 두 행에서 CV, AIC, 및 AICc 값이 거의 같습니다.
- 우리는 생산 변수를 무시할 수 있고 비슷한 예측치를 얻을 수 있었습니다.
- 생산과 실업률이 아주 높은 상관 관계를 갖고, 따라서 생산에 나타나는 대부분의 예측 정보는 실업률 변수에도 포함된다는 것에 주목하시길 바랍니다.

소득	생산	저축	실업률	CV	AIC	AICc	BIC	AdjR2
1	1	1	1	0.116	-409.3	-408.8	-389.9	0.749
1	0	1	1	0.116	-408.1	-407.8	-391.9	0.746
1	1	1	0	0.118	-407.5	-407.1	-391.3	0.745
1	0	1	0	0.129	-388.7	-388.5	-375.8	0.716
1	1	0	1	0.278	-243.2	-242.8	-227.0	0.386
1	0	0	1	0.283	-237.9	-237.7	-225.0	0.365
1	1	0	0	0.289	-236.1	-235.9	-223.2	0.359
0	1	1	1	0.293	-234.4	-234.0	-218.2	0.356
0	1	1	0	0.300	-228.9	-228.7	-216.0	0.334
0	1	0	1	0.303	-226.3	-226.1	-213.4	0.324
0	0	1	1	0.306	-224.6	-224.4	-211.7	0.318
0	1	0	0	0.314	-219.6	-219.5	-209.9	0.296
0	0	0	1	0.314	-217.7	-217.5	-208.0	0.288
1	0	0	0	0.372	-185.4	-185.3	-175.7	0.154
0	0	1	0	0.414	-164.1	-164.0	-154.4	0.052
0	0	0	0	0.432	-155.1	-155.0	-148.6	0.000



4.5 회귀로 예측하기

- 회귀로 예측

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1,t} + \hat{\beta}_2 x_{2,t} + \cdots + \hat{\beta}_k x_{k,t},$$

- 여기에서 추정된 계수를 포함하고 회귀식의 오차를 무시합니다. 예측 변수(predictor variable)의 값들 $x_{1,t}, \dots, x_{k,t}$ for $t = 1, \dots, T$ 을 넣으면 y 의 (학습-표본) 적합값을 얻습니다. 하지만, 여기에서는 y 의 미래값을 예측하는 것에 관심이 있습니다.

4.5 회귀로 예측하기

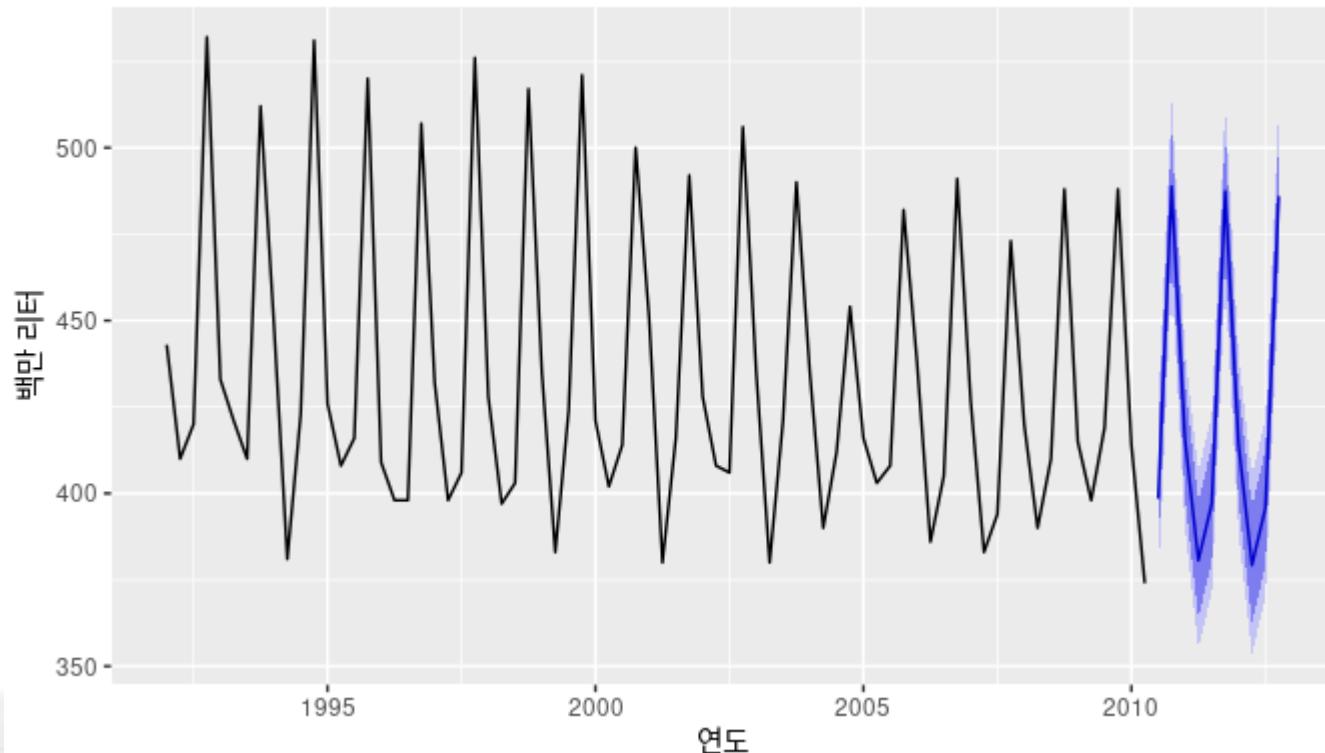
- 사전 예측 대 사후 예측
 - 사전 예상값(Ex-ante forecast)
 - 미리 쓸 수 있는 정보만으로 내는 것입니다.
 - 예를 들어, 표본이 끝난 다음 분기의 미국 소비 백분율 변화에 대한 사전 예상값에는 2016 Q3까지 사용할 수 있는 정보만 사용해야 합니다.
 - 이는 당시에 이용할 수 있는 정보를 가지고 미리 낸 진정한 예측값입니다.
 - 따라서 사전 예상값을 내려면 모델에 예측변수(predictor variable)의 미래 값(예상값)이 필요합니다.
 - 사후 예상값(Ex-post forecast)
 - 예측변수(predictor variable)에 대한 이후 정보로 내는 것입니다.
 - 예를 들어, 소비에 대한 사후 예상값을 낼 때 일단 예측변수(predictor variable)의 실제 값을 관측하고나서 실제 값을 사용할 수도 있습니다.
 - 이러한 예상값은 실제 예상값은 아니지만, 예측 모델이 작동하는 것을 살펴볼 때 유용합니다.
 - 사후 예상값을 내는 모델을 예측 기간의 데이터를 이용하여 추정해서는 안 됩니다.
 - 즉, 사후 예상값은 예측변수(x 변수)에 관한 지식을 가정할 수 있지만, 예측할 데이터(y 변수)에 대한 지식을 가정해서는 안 됩니다.

4.5 회귀로 예측하기

- 예제: 호주 분기별 맥주 생산량

- 사전 예상값을 낼 때 예측변수(predictor variable)의 실제 미래 값을 쓸 수 없습니다. 왜냐하면 예측변수(predictor variable)의 값을 미리 알 수 없기 때문입니다.

회귀를 이용한 맥주 생산량 예측값

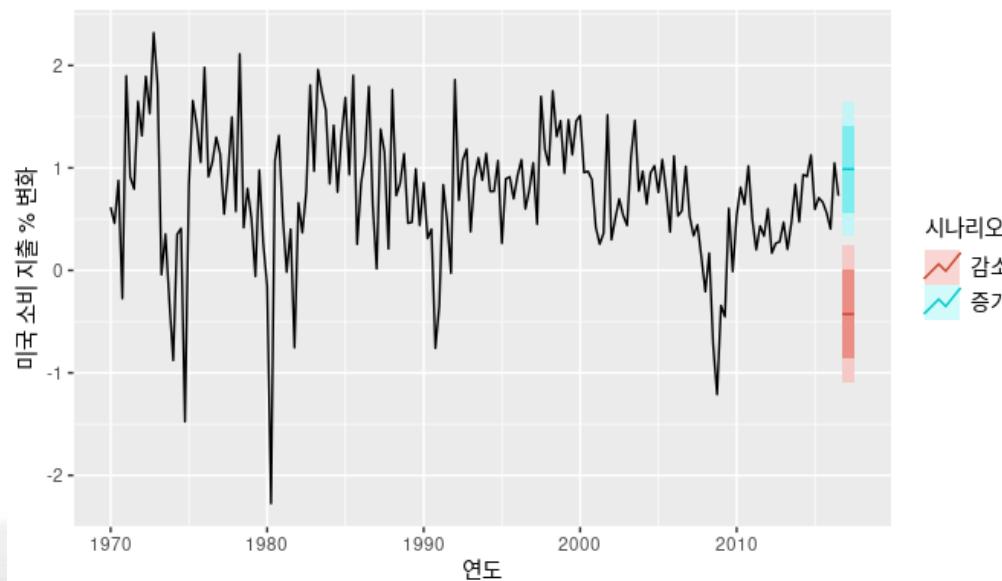


맥주 생산량에 대한 회귀 모델로부터 얻은 예측값. 상대적으로 어두운 음영 영역은 80% 예측구간(prediction interval)이고 상대적으로 밝은 음영 영역은 95% 예측구간(prediction interval)입니다.

4.5 회귀로 예측하기

• 시나리오 기반 예측

- 예측하는 사람은 관심 있는 예측변수(predictor variable)에 대한 가능한 시나리오를 가정합니다.
 - 예를 들어, 미국 정책 담당자는 고용률 변화 없이 소득과 저축에 대해 각각 1%와 0.5%의 지속적인 성장이 일어나는 것과, 각각 1%와 0.5%의 감소가 일어나는 상황에서 소비에서 예측된 변화량을 표본 이후 각 네 분기에 대해 비교하는 일에 관심이 있을 수 있습니다.



시나리오 기반 예측으로 미국 개인 소비 지출 백분율 변화 예측하기.

- 시나리오 기반의 예측치에 대해서는 예측 구간(prediction interval)이 예측 변수(predictor variable)의 미래 값과 관련된 불확실성을 포함하지 않습니다. 예측 변수(predictor variable)의 값을 미리 있다고 가정

4.5 회귀로 예측하기

- 예측 회귀 모델 세우기
 - 회귀 모델의 가장 큰 장점은 관심 있는 목표 예상변수(forecast variable)와 예측변수(predictor variable)의 중요한 관계를 잡아내는데 사용할 수 있다는 것입니다.
 - 하지만, 주된 어려움은 사전 예상값(ex-ante forecast)을 내기 위해, 모델에 각 예측변수(predictor variable)의 미래값이 필요하다는 것입니다. 시나리오 기반 예측에 관심이 있다면, 이러한 모델은 유용합니다.
 - 하지만, 사후 예측에 주된 관심이 있다면, 예측변수(predictor variable)에 대한 예측값을 얻는 것은 어려운 일이 될 수 있습니다(예측변수에 대한 예측값을 내는 많은 경우는 예측변수(predictor variable) 없이 목적 예상변수(forecast variable)를 직접 예측하는 것보다 더 어려울 수 있습니다).

4.5 회귀로 예측하기

• 예측 회귀 모델 세우기

- 대안은 예측변수(predictor variable)로 시차 값(lagged value)을 사용하는 것입니다. h -단계 앞 예측을 내는데 관심이 있다고 하면 $h = 1, 2, \dots$ 에 대해 다음과 같이 쓸 수 있습니다.

$$y_{t+h} = \beta_0 + \beta_1 x_{1,t} + \cdots + \beta_k x_{k,t} + \varepsilon_{t+h}$$

- 예측변수(predictor variable) 모음은 y 를 관찰하기 전 h 시간 주기에 관찰한 x 의 값으로 구성됩니다.
- 따라서, 추정된 모델을 미래에 투영할 때, (즉, T 표본 넘어), 모든 예측변수(predictor variable) 값을 이용할 수 있습니다.
- 예측변수(predictor variable)의 시차값(lagged value)을 포함하면 예측값을 쉽게 낼 수 있도록 모델을 작동시키기도 하면서, 직관적으로 이해할 수 있도록 만들기도 합니다.
 - 예를 들면, 생산 증가를 목표로 하는 정책 변화의 효과는 소비 지출에 바로 영향을 미치지 않을 수도 있습니다.

4.5 회귀로 예측하기

• 예측구간

- 다중 회귀 모델에 대한 예측구간(prediction interval)을 계산하는 방법에 대한 일반적인 식을 다룹니다. 여기에는 몇 가지 고급 행렬 연산이 들어가기 때문에 여기에서는 간단한 회귀에 대한 예측구간(prediction interval)을 계산하는 경우만 다룹니다. 여기에서 한 예측값을 다음과 같은 식으로 낼 수 있습니다,

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x.$$

- 회귀 오차가 정규 분포를 따른다는 것을 가정하면, 이 예측과 관련된 근사적인 95% 예측구간(예측 간격이라고도 불립니다)은 다음과 같이 주어집니다.

$$\hat{y} \pm 1.96 \hat{\sigma}_e \sqrt{1 + \frac{1}{T} + \frac{(x - \bar{x})^2}{(T - 1)s_x^2}}, \quad \hat{\sigma}_e = \sqrt{\frac{1}{T - k - 1} \sum_{t=1}^T e_t^2}.$$

T 는 관측값의 전체 개수, \bar{x} 는 관측된 x 의 평균, s_x 는 관측된 x 값의 표준 편차입니다. 비슷하게, 80% 예측 구간은 1.96을 1.28로 바꾸어 얻을 수 있습니다.

- 다중 회귀 모델

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_k x_{k,t} + \varepsilon_t$$

- 여기에서 ε_t 은 평균이 0이고 분산이 σ^2 입니다. 이것은 목표 예상 변수(forecast variable)와 예측변수(predictor variable) 사이의 관계를 나타냅니다.
- 행렬 형태로 쓰면...

$$\begin{aligned} \mathbf{y} &= (y_1, \dots, y_T)', \quad \boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_T)', \\ \boldsymbol{\beta} &= (\beta_0, \dots, \beta_k)' \end{aligned} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & x_{2,1} & \cdots & x_{k,1} \\ 1 & x_{1,2} & x_{2,2} & \cdots & x_{k,2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1,T} & x_{2,T} & \cdots & x_{k,T} \end{bmatrix}.$$

- 그러면 아래와 같은 식을 얻습니다.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

여기에서 $\boldsymbol{\varepsilon}$ 은 평균이 0이고 분산이 $\sigma^2\mathbf{I}$ 입니다. 행렬 \mathbf{X} 에는 관측값의 수를 나타내는 T 행이 있고, 예측변수(predictor variable)의 수에 1을 더한 열로 표현되는 절편을 나타내는 $k+1$ 열이 있습니다.

- 최소 제곱 추정

- 다음과 같은 표현식을 최소화하여 최소 제곱 추정치를 얻습니다. $\varepsilon' \varepsilon = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$. β 가 아래와 같은 값을 가질 때...

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

- 위의 표현식이 최소가 되는 것을 보일 수 있습니다. 이것은 때때로 “정규식(normal equation)”이라고 알려져 있습니다.
- 추정된 계수에는 역행렬 $\mathbf{X}'\mathbf{X}$ 이 필요합니다. 만약에 \mathbf{X} 가 전체 열 순위(full column rank)가 아니면, 행렬 $\mathbf{X}'\mathbf{X}$ 은 단수(singular)이며 모델을 추정할 수 없습니다.
- 아래의 식을 이용하여 잔차(residual) 분산을 추정합니다.

$$\hat{\sigma}_e^2 = \frac{1}{T - k - 1} (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta}).$$

4.6 행렬 정식화

- 적합값과 교차검증

- 정규식은

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{H}\mathbf{y},$$

- 을 이용하여, 적합값을 계산할 수 있다는 것을 나타냅니다.
 - 여기에서 $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ 는 $\hat{\mathbf{y}}$ (“y-모자(y-hat)”)을 계산할 때 사용되기 때문에 “모자-행렬(hat-matrix)”로 알려져 있습니다.
 - \mathbf{H} 의 대각 성분 값을 h_1, \dots, h_T 로 쓰면, 아래와 같이 교차검증 (cross-validation) 통계를 계산할 수 있습니다.

$$CV = \frac{1}{T} \sum_{t=1}^T [e_t / (1 - h_t)]^2,$$

- 여기에서 e_t 는 T 개의 모든 관측값에 모델을 맞춘 것으로부터 얻은 잔차입니다. 따라서, CV 통계를 계산할 때 T 개의 나뉜 모델을 실제로 맞출 필요가 없습니다.

- 예측값과 예측구간
 - x^* 을 예측값을 내려는 예측변수(predictor variable)의 값을 포함하는 행 벡터로 둡시다(X 와 같은 형태). 그러면 예측값은 다음과 같이

$$\hat{y} = \mathbf{x}^* \hat{\beta} = \mathbf{x}^* (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}$$

- 로 주어지고, 추정된 분산은

$$\hat{\sigma}_e^2 [1 + \mathbf{x}^* (\mathbf{X}' \mathbf{X})^{-1} (\mathbf{x}^*)'] .$$

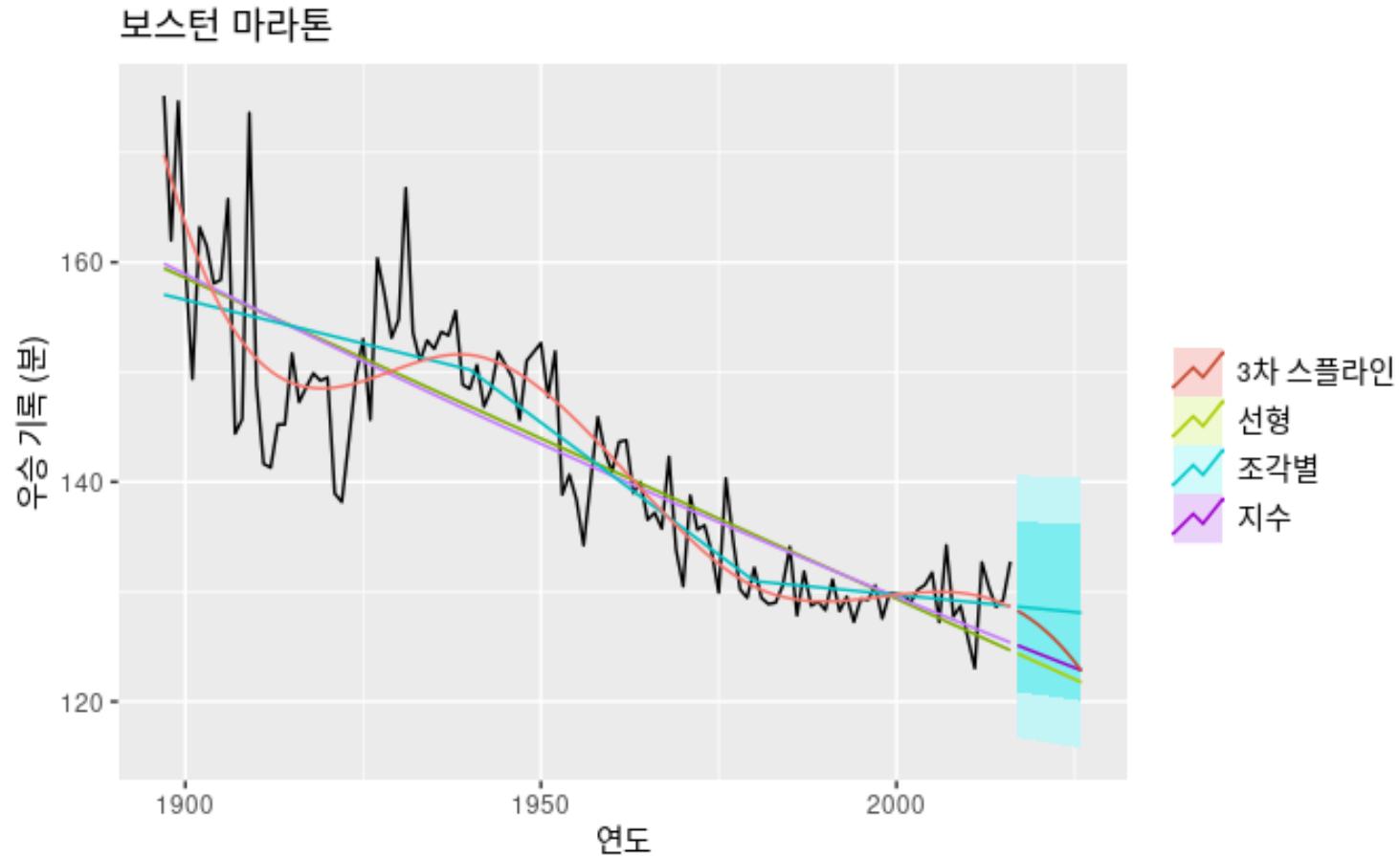
- 로 주어집니다. 95% 예측구간은 (오차가 정규 분포를 따른다고 가정하고) 다음과 같이

$$\hat{y} \pm 1.96 \hat{\sigma}_e \sqrt{1 + \mathbf{x}^* (\mathbf{X}' \mathbf{X})^{-1} (\mathbf{x}^*)'}.$$

- 로 계산할 수 있습니다. 이 식은 오차 항 ε 때문에 생기는 불확실성과 계수 추정값의 불확실성을 고려합니다. 하지만, x^* 에 있는 모든 오차를 무시합니다.
- 따라서 예측변수의 미래값이 확실하지 않을 때, 이 표현식으로 계산한 예측구간은 너무 좁을 것입니다.

4.7 비선형 회귀

- 예제: 보스턴 마라톤 우승 기록



- 상관관계는 인과관계가 아닙니다
 - 상관관계(correlation)를 인과관계(causation)와 또는 인과관계를 예측하는 것과 혼동하지 않는 것은 중요합니다.
 - 변수 x 는 변수 y 를 예측할 때 유용할 수 있습니다만, 이것이 x 가 y 의 원인이 된다는 의미는 아닙니다. x 가 y 의 원인이 될 수 있습니까만, y 가 x 의 원인이 될 수도 있고, 또는 둘 사이의 관계는 단순한 인과관계보다 더 복잡할 수 있습니다.

- 다중공선성과 예측
 - 다중 회귀에 2개 이상의 예측변수(predictor variable)로 비슷한 정보가 주어질 때 나타나는 **다중공선성(Multicollinearity)**입니다.
 - 이 성질은 두 예측변수(predictor variable)의 상관관계가 아주 높은 상황(즉, 상관 계수가 1이나 -1에 가까울 때)에서 나타날 수 있습니다. 이 경우에, 변수 중에서 하나의 값을 알면 다른 변수의 값을 추정할 수 있습니다. 따라서, 두 예측변수는 비슷한 정보를 주고 있는 셈입니다.
 - 예를 들면, 발 크기는 키를 예측하는데 사용할 수 있습니다만, 같은 모델에서 왼쪽 발과 오른쪽 발의 크기를 모두 사용한다고 해서 예측값을 더 나쁘게 만들지는 않을 것이더라도 더 나아지게 만들지는 않을 것입니다.
 - 상관관계가 크면(+1이나 -1에 가깝지만 같지는 않은), 회귀 계수를 계산하여 추정하는 것이 어렵습니다. 사실, 몇몇 소프트웨어(특히, 마이크로소프트 엑셀)은 아주 정확하지 않은 계수 추정치를 낼 수 있습니다.

- 다중공선성과 예측
 - 다중공선성이 나타날 때, 각 회귀 계수와 관련된 불확실성은 클 것입니다. 이것은 추정하기 어렵기 때문입니다.
 - 결과적으로, 회귀 계수에 대한 통계 검정(즉, t-검정)을 신뢰할 수 없게 됩니다. (예측할 때는 이러한 검정에 거의 관심을 두지 않습니다.) 또한, 예측값으로 이어지는 각각의 나눈 예측변수에 대해 정확하게 서술할 수 없을 것입니다.
 - 좋은 통계 소프트웨어를 사용하고 있고, 예측변수(predictor variable) 각각이 구체적으로 기여하는 양에 관심이 없고, 예측변수의 미래 값이 과거 범위에 들어가면, 특별히 걱정할 필요가 없습니다 — 완벽한 상관관계가 나타나는 경우를 제외하고는 다중공선성(multicollinearity)이 문제가 되지 않습니다.

5.1 시계열 분해

- 시계열 분해
 - 시계열 데이터는 다양한 패턴으로 나타날 수 있습니다. 시계열을 몇 가지 성분으로 나누려는 작업은 시계열을 이해하는데 종종 도움이 됩니다. 이러한 각각의 성분은 기본적인 패턴 범주를 나타냅니다.
 - 추세(trend), 계절성(seasonality), 주기(cycle) 이렇게 세 가지 시계열 패턴을 살펴봤습니다.
 - 시계열을 이러한 성분으로 나눌 때, 종종 추세와 주기를 결합하여 하나의 **추세-주기** 성분으로 다룹니다 (때때로 단순히 **추세**라고 부르기도 합니다).
 - 그래서 시계열이 다음과 같이 세 가지 성분으로 구성된다고 생각합니다: 추세-주기 성분, 계절성 성분, (시계열의 나머지 요소를 포함하는) 나머지(remainder) 성분.
 - 이 장에서는 시계열에서 위와 같은 성분을 추출할 때 사용하는 몇 가지 일반적인 기법을 다룹니다. 시계열을 더 잘 이해하기 위해 종종 이러한 작업을 진행하고, 이 작업의 결과는 예측 정확도를 높이는데 사용될 수 있습니다.

5.1 시계열 성분

- 덧셈 분해(additive decomposition)

$$y_t = S_t + T_t + R_t$$

- y_t 는 데이터이고, S_t 는 계절성분, T_t 는 추세-주기 성분, R_t 는 나머지 성분입니다. y_t, S_t, T_t, R_t 는 시점 t 에서의 양입니다.
- 계절성 요동의 크기나 추세-주기 주위의 변동이 시계열의 수준에 의해 변하지 않으면, 덧셈 분해가 가장 적절합니다

- 곱셈 분해(multiplicative decomposition)

$$y_t = S_t \times T_t \times R_t \quad \log y_t = \log S_t + \log T_t + \log R_t$$

- 계절성 패턴에서 변동이나 추세-주기 주위의 변동이 시계열의 수준(level)에 비례하는 것으로 나타날 때는, 곱셈 분해가 더 적절합니다. 흔히 경제 분야 시계열을 다룰 때 곱셈 분해를 사용합니다.

5.1 시계열 성분

- 전자 장비 제조 예시

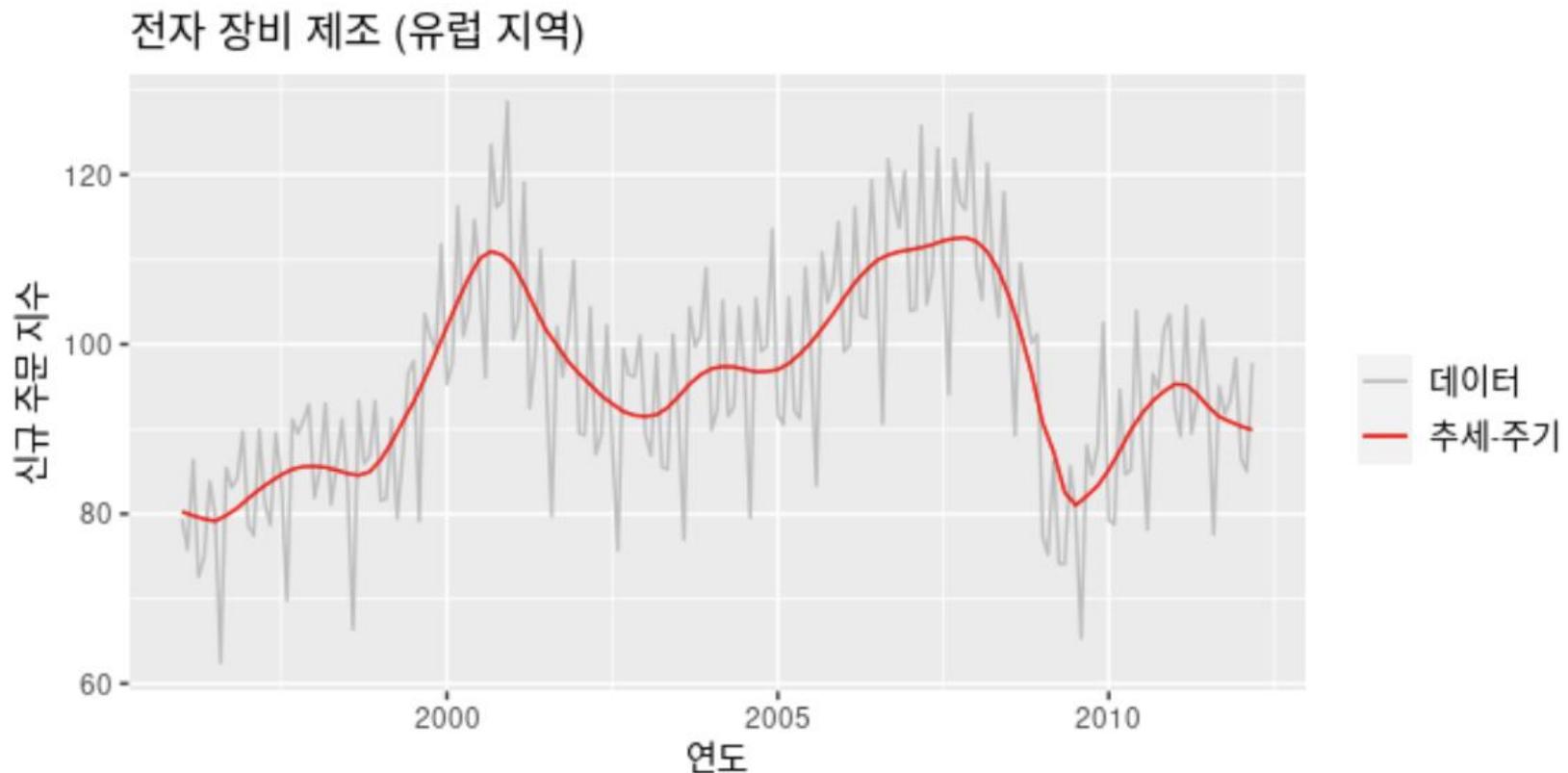


Figure 6.1: 전자 장비 주문: 추세-주기 성분 (빨간색) 그리고 원본 데이터 (회색).

5.1 시계열 성분

- 전자 장비 제조 (덧셈 분해 예시; STL 기법 사용)

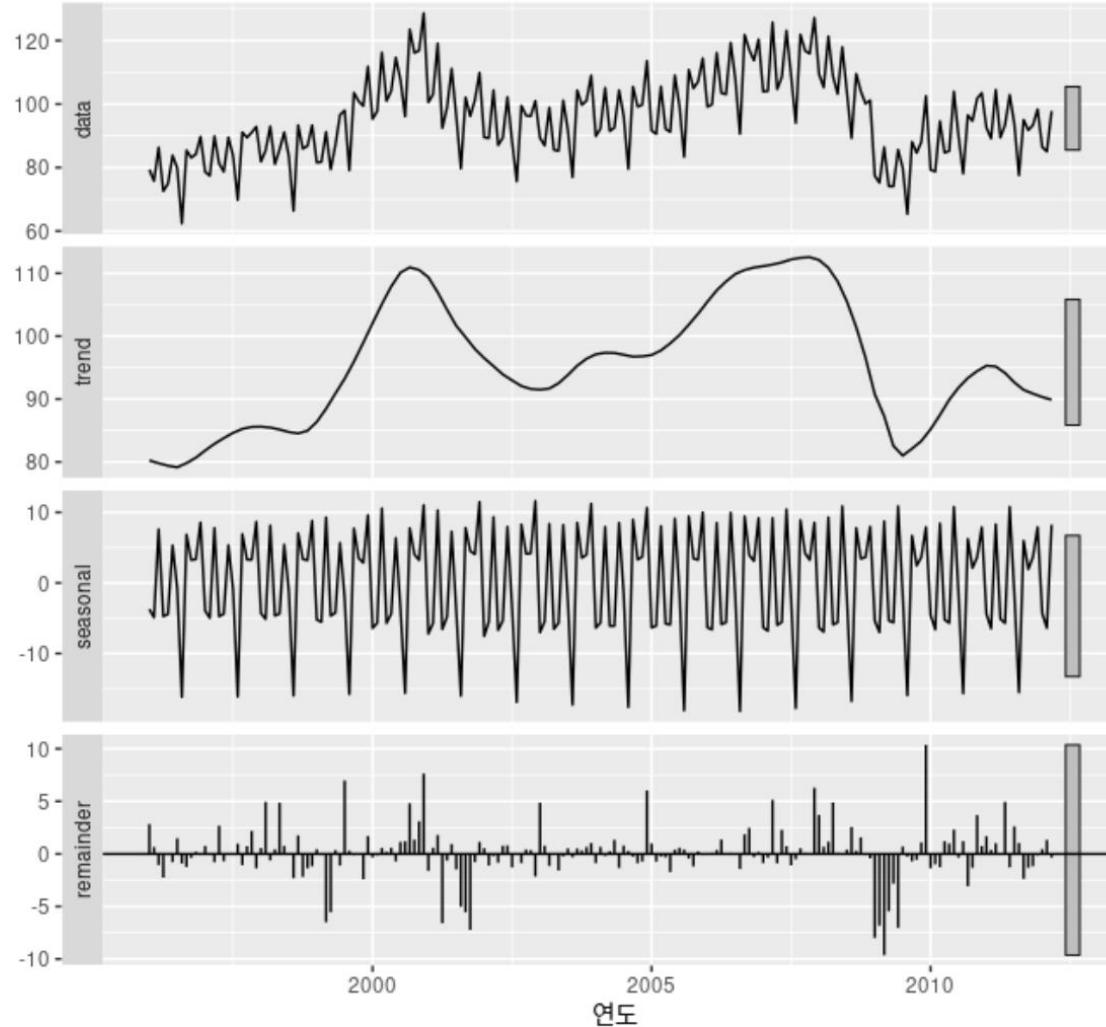
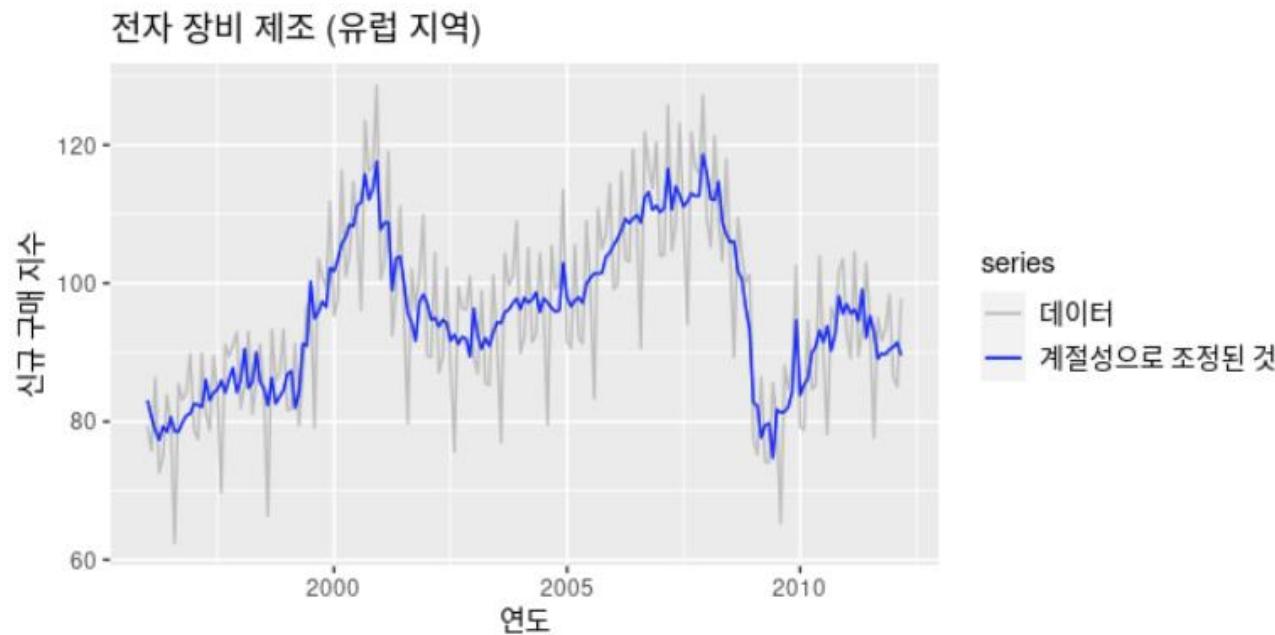


Figure 6.2: 전자 장비 주문(위)과 세 가지 덧셈 성분.

5.1 시계열 성분

- 계절성으로 조정된 데이터

- 계절성으로 조정된(seasonally adjusted)" 데이터라고 부릅니다. 덧셈 분해에서, 계절성으로 조정된 데이터는 $y_t - S_t = T_t + R_t$ 로 주어지고, 곱셈 분해에서는 $y_t/S_t = T_t \times R_t$ 을 이용하여 계절성으로 조정된 값



5.2 이동평균

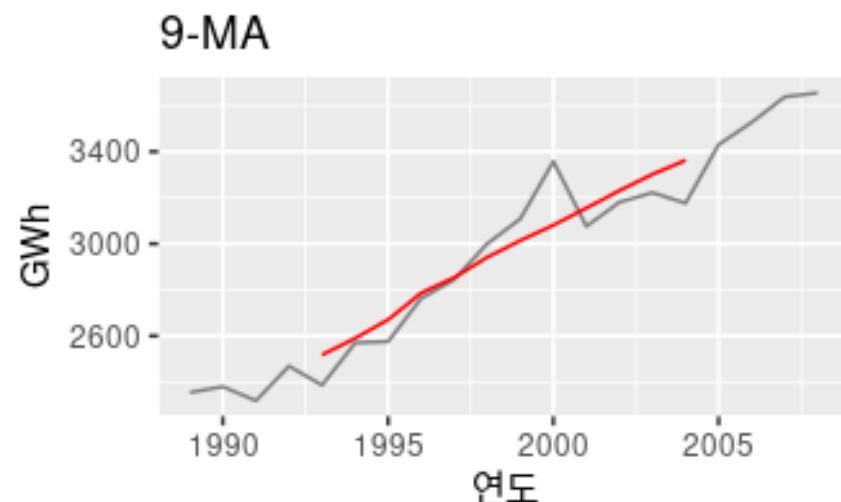
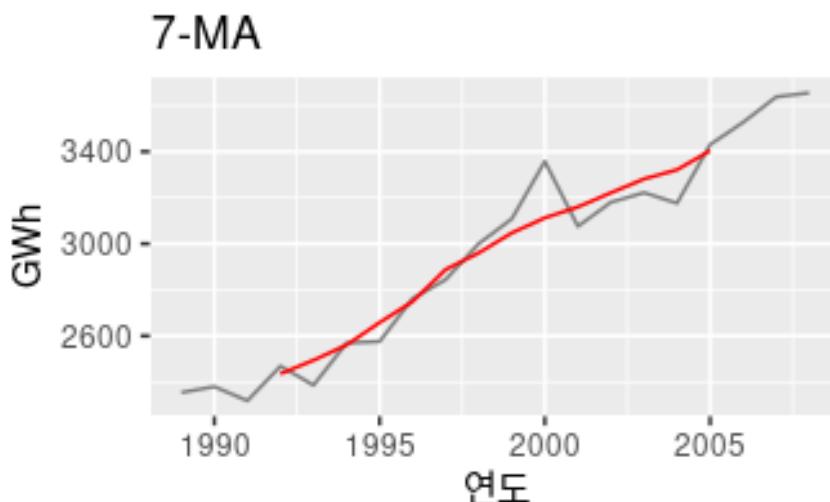
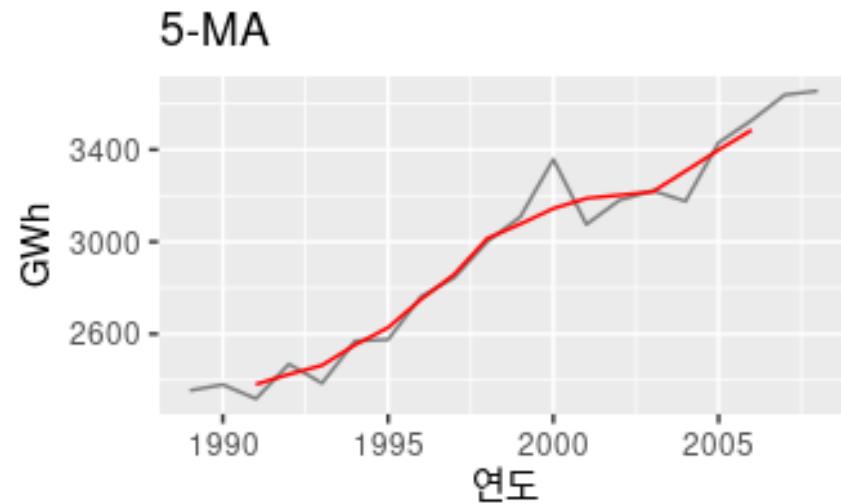
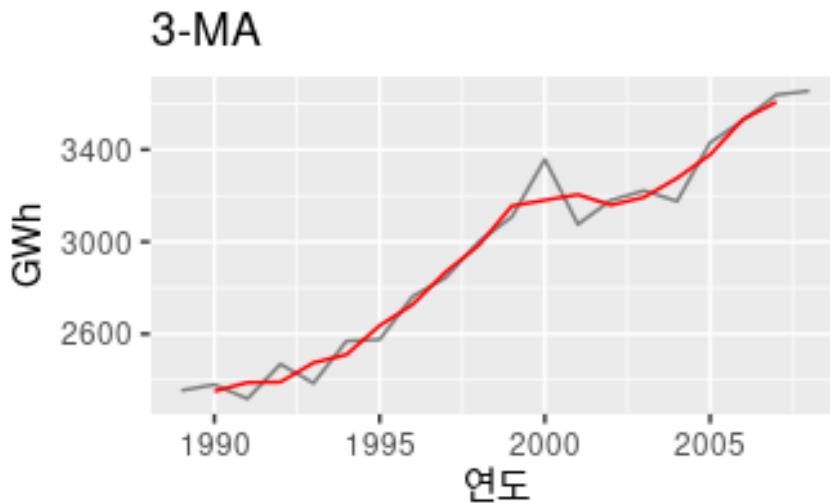
- 이동평균 (moving average)
 - 고전적인 시계열 분해 기법은 1920년까지 거슬러 올라가며, 1950년대까지 널리 사용
 - 이 방법은 후대의 시계열 분석 기법의 기초가 되고 있기 때문에, 이 방법이 어떻게 작동하는지 잘 이해하는 것은 중요합니다.
 - 고전적인 분해 방법의 첫 번째 단계는 추세-주기를 측정하기 위해 이동평균(moving average) 방법을 사용
- 이동 평균 평활
 - 차수(order) m 의 이동 평균은 다음과 같이 쓸 수 있습니다.

$$\hat{T}_t = \frac{1}{m} \sum_{j=-k}^k y_{t+j},$$

여기에서 $m = 2k + 1$ 입니다. 즉, k 기간 안의 시계열 값을 평균하여 시간 t 의 추세-주기를 측정합니다. 측정 시기가 비슷하면 값도 비슷하기도 합니다. 따라서 평균이 데이터의 무작위성을 줄이고 매끄러운 추세-주기 성분만 남깁니다. 이것을 차수 m 의 이동 평균이라는 의미에서 " $m - MA$ "라고 부릅니다

5.2 이동평균

- 다른 이동 평균을 주거용 전력 판매량 데이터에 적용



5.3 고전적인 분해법

- 고전적인 분해에 대한 문제점
 - 처음 몇 개와 마지막 몇 개의 관측값에 대한 추세 추정값을 얻을 수 없습니다. 예를 들면, $m = 12$ 일 때, 처음 여섯개와 마지막 여섯개 관측값에 대한 추세 추정값은 존재하지 않습니다. 결과적으로, 같은 기간에 대해 나머지 성분을 측정하는 것도 불가능합니다.
 - 추세-주기 측정은 데이터에 나타나는 급격한 증가나 감소를 과도하게 매끄럽게 합니다.
 - 고전적인 분해법은 계절성분이 매년 반복된다는 것을 가정합니다. 많은 시계열에서, 이렇게 가정하는 것이 나름 합리적입니다만, 몇몇 더 긴 시계열에 대해서는 그렇지 않습니다.
 - 예를 들면, 전기 수요 패턴은 에어컨이 더욱 보급되면서 시간에 따라 변했습니다. 따라서 많은 지역에서 현재 계절적인 패턴에서 최대 수요 값이 에어컨 때문에 여름에 나타나지만, 수십년 전에는 난방 때문에 겨울에 나타났습니다. 고전적인 분해 기법으로는 이러한 시간적으로 변하는 계절적인 변화를 다룰 수 없습니다.
 - 가끔 짧은 기간의 시계열 값이 특별히 이상할 수 있습니다.
 - 예를 들면, 월별 항공 탑승자 수는 노사분규에 영향을 받을 수 있고 분쟁 기간 동안 평소보다 특이한 패턴이 나타날 수 있습니다. 고전적인 방법은 이러한 종류의 특이한 값을 다루기에 적절하지 않습니다.

- 분해법 종류 (Dagum & Bianconcini 2016)
 - X11 분해
 - 분기별 데이터와 월별 데이터를 분해할 때 인기 있는 또 하나의 기법은 미국 인구 조사국(the US Census Bureau)과 캐나다 통계청(Statistics Canada)에서 창안한 X11 기법입니다.
 - SEATS 분해
 - Seasonal Extraction in ARIMA Time Series (SEATS)의 약자로 분기별 데이터와 월별 데이터에서만 작동합니다. 따라서 일별, 시간별, 주별 등 다른 종류의 계절성은 다른 접근 방식으로 다뤄야 합니다.
 - STL 분해
 - Seasonal and Trend decomposition using Loess (STL)의 약자입니다. 여기에서 Loess는 비선형 관계를 추정하기 위한 기법입니다
 - SEAT와 X11과는 다르게, STL은 월별이나 분기별 데이터를 포함하여 어떤 종류의 계절성도 다룰 수 있습니다.
 - 계절적인 성분이 시간에 따라 변해도 괜찮습니다. 계절성분의 변화율을 사용자가 조절할 수 있습니다.
 - 추세-주기의 매끄러운 정도를 사용자가 조절할 수 있습니다.
 - 가끔 있는 이상값이 추세-주기와 계절성분에 영향을 주지 않게 만들 수 있습니다 (즉, 사용자가 강력한 분해법을 명시할 수 있습니다). 하지만, 이상값은 나머지 성분(remainder)에 영향을 줄 것입니다
 - 반면에, STL은 몇 가지 단점을 가지고 있습니다. 특별히, **거래일이나 달력 변동을 자동으로 다루지 않고, 덧셈 분해만 지원합니다**

5.5 추세와 계절성의 강도를 측정하기

- 추세 강도 측정

$$y_t = S_t + T_t + R_t$$

$$F_T = \max \left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(T_t + R_t)} \right). \quad F_S = \max \left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(S_t + R_t)} \right).$$

- 추세가 강하게 나타나는 데이터에서는, 계절성으로 조정된 데이터가 나머지 성분보다 훨씬 더 큰 변동성을 나타내야 합니다. $\text{Var}(R_t)/\text{Var}(T_t + R_t)$ 가 상대적으로 작아야 합니다.
- 하지만 추세가 거의 없거나 아예 없는 데이터의 경우에는, 두 분산값이 근사적으로 같아야 합니다. 따라서 추세의 강도를 다음과 같이 정의합니다
- 이 식을 통해 추세의 강도를 0과 1 사이 값으로 얻을 수 있습니다. 나머지 성분의 분산이 계절성으로 조정된 데이터의 분산보다 훨씬 더 클 수 있기 때문에, F_t 의 가능한 최소값을 0으로 두었습니다.
- 시계열의 계절성 강도 F_s 가 0에 가까우면 거의 계절성이 없다는 것을 의미하고, $\text{Var}(R_t)$ 이 $\text{Var}(S_t + R_t)$ 에 비해 훨씬 작을 것이기 때문에 시계열에서 계절성이 강하게 나타나면 F_s 가 1에 가깝게 나타날 것입니다.

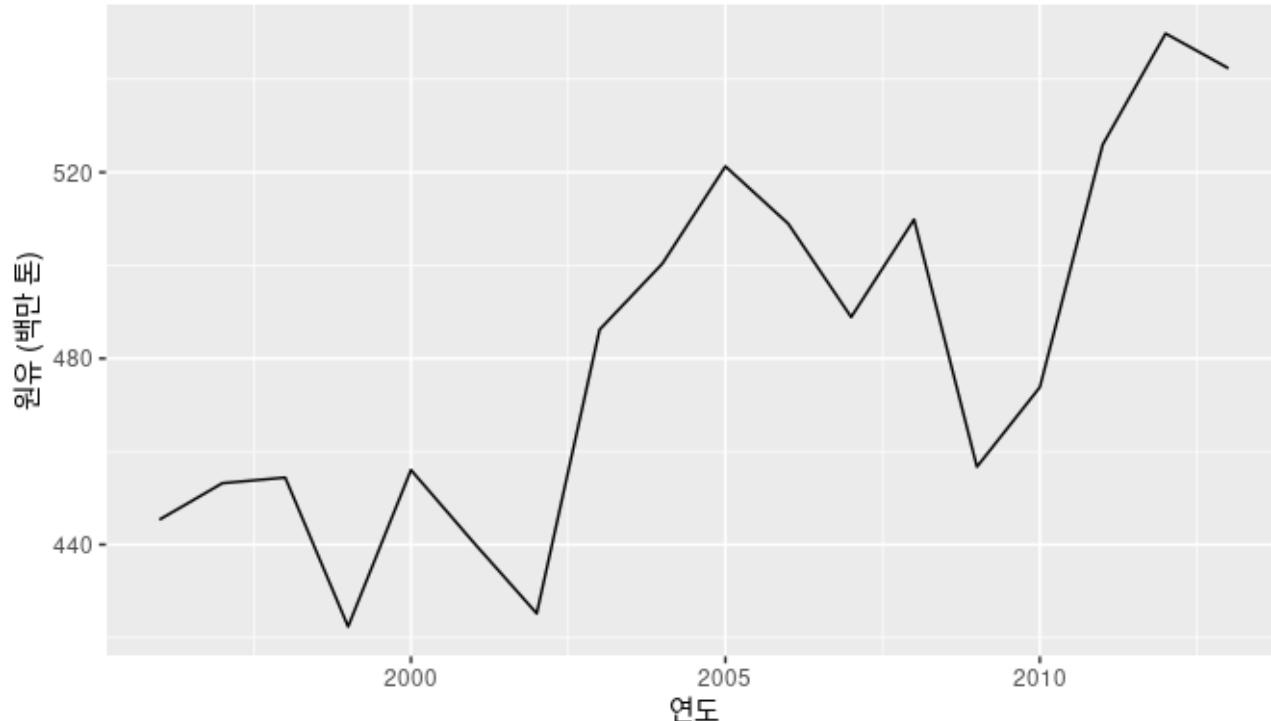
시계열 데이터를 이용한 지수 평활

- 지수평활

- 지수평활(exponential smoothing)은 1950년대 후반에 제안되었고 ([Brown, 1959](#); [Holt, 1957](#); [Winters, 1960](#)), 가장 성공적인 몇 가지 예측 기법에 영향을 주었습니다.
- 지수 평활을 사용하여 얻은 예측값은 과거 관측값의 가중평균 (weighted average)입니다. 여기에서 과거 관측값은 오래될 수록 지수적으로 감소하는 가중치를 갖습니다.
- 다르게 말하면, 가장 최근 관측값이 가장 높은 가중치를 갖는다는 말입니다. 이러한 방식으로 다양한 종류의 시계열을 가지고 신뢰할만한 예측 작업을 빠르게 수행할 수 있다는 것은 엄청난 장점이고, 이는 산업 분야에 응용할 때 아주 중요한 부분입니다.

6.1 단순 지수 평활

- 단순 지수평활 (simple exponential smoothing, SES)
 - 추세나 계절성 패턴이 없는 데이터를 예측할 때 쓰기 좋습니다



예를 들면, 위에 있는 데이터를 보면 평균값이 시간에 따라 느리게 변하는 것 같긴 하지만, 분명한 추세나 계절성이 보이지 않습니다. (마지막 몇 년 동안 증가하는 사실이 추세가 있다는 것을 시사합니다).

6.1 단순 지수 평활

- 단순 지수평활 (simple exponential smoothing, SES)
 - 오래된 관측값보다 더 최근 관측값에 더 큰 가중치를 주는 경우

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1 - \alpha)y_{T-1} + \alpha(1 - \alpha)^2 y_{T-2} + \dots,$$

- 여기에서 $0 \leq \alpha \leq 1$ 은 평활 매개변수입니다. 시간 $T + 1$ 에 대한 한 단계 앞 예측치(one-step-ahead forecast)는 시계열 y_1, \dots, y_T 에서 모든 관측값을 가중 평균하여 얻은 값입니다. 가중치가 감소하는 비율은 매개변수 α 로 조절합니다

	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 0.8$
y_T	0.2000	0.4000	0.6000	0.8000
y_{T-1}	0.1600	0.2400	0.2400	0.1600
y_{T-2}	0.1280	0.1440	0.0960	0.0320
y_{T-3}	0.1024	0.0864	0.0384	0.0064
y_{T-4}	0.0819	0.0518	0.0154	0.0013
y_{T-5}	0.0655	0.0311	0.0061	0.0003

단순 지수 평활로 예측할 때, 4개의 서로 다른 α 로 관측값에 가중치를 할당한 것을 나타냅니다. 작은 α 값이라도 가중치의 합은 어떠한 적절한 표본 크기에 대해서도 근사적으로 1이 될 것입니다.

6.1 단순 지수 평활

- 가중 평균 형태

- 시간 $T + 1$ 의 예측은 가장 최근 관측값 y_T 와 이전 예측값 $\hat{y}_{T|T-1}$ 의 가중평균과 같습니다:

$$\hat{y}_{T+1|t} = \alpha y_T + (1 - \alpha) \hat{y}_{T|T-1}$$

- 여기에서 $0 \leq \alpha \leq 1$ 은 평활 매개변수입니다. 비슷하게, $t = 1, \dots, T$ 에 대해, 적합값(fitted value)도 다음과 같이 적을 수 있습니다.

$$\hat{y}_{t+1|t} = \alpha y_t + (1 - \alpha) \hat{y}_{t|t-1},$$

(여기에서 적합값은 단순히 학습 데이터의 한 단계 예측이라는 것을 기억하시면 됩니다.)

6.1 단순 지수 평활

- 가중 평균 형태

- 진행 과정을 어디에선가 시작해야만 하니, (우리가 추정해야할) 시간 1에서의 첫 번째 적합값을 ℓ_0 로 둡시다. 그러면,

$$\begin{aligned}\hat{y}_{2|1} &= \alpha y_1 + (1 - \alpha)\ell_0 \\ \hat{y}_{3|2} &= \alpha y_2 + (1 - \alpha)\hat{y}_{2|1} \\ \hat{y}_{4|3} &= \alpha y_3 + (1 - \alpha)\hat{y}_{3|2} \\ &\vdots \\ \hat{y}_{T|T-1} &= \alpha y_{T-1} + (1 - \alpha)\hat{y}_{T-1|T-2} \\ \hat{y}_{T+1|T} &= \alpha y_T + (1 - \alpha)\hat{y}_{T|T-1}.\end{aligned}$$

각 식을 다음의 식에 대입하면 오른쪽의 결과를 얻습니다.

$$\begin{aligned}\hat{y}_{3|2} &= \alpha y_2 + (1 - \alpha)[\alpha y_1 + (1 - \alpha)\ell_0] \\ &= \alpha y_2 + \alpha(1 - \alpha)y_1 + (1 - \alpha)^2\ell_0 \\ \hat{y}_{4|3} &= \alpha y_3 + (1 - \alpha)[\alpha y_2 + \alpha(1 - \alpha)y_1 + (1 - \alpha)^2\ell_0] \\ &= \alpha y_3 + \alpha(1 - \alpha)y_2 + \alpha(1 - \alpha)^2y_1 + (1 - \alpha)^3\ell_0 \\ &\vdots \\ \hat{y}_{T+1|T} &= \sum_{j=0}^{T-1} \alpha(1 - \alpha)^j y_{T-j} + (1 - \alpha)^T \ell_0.\end{aligned}$$

마지막 항은 T 값이 클 때 작아집니다. 그래서 가중 평균 형태는 같은 예측식이 됩니다.

6.1 단순 지수 평활

- 성분 형태

Forecast equation

$$\hat{y}_{t+h|t} = \ell_t$$

Smoothing equation

$$\ell_t = \alpha y_t + (1 - \alpha) \ell_{t-1},$$

- 여기서 ℓ_t 는 시간 t 에서 시계열의 수준값(또는 평활화된 값)입니다. $h = 1$ 로 두면 적합값을 얻을 수 있고, $t = T$ 로 두면 학습 데이터 이후의 예측값을 얻을 수 있습니다.
- 예측식은 시간 $t + 1$ 의 예측값은 시간 t 의 추정된 수준값입니다. (종종 수준식이라고 부르는) 수준에 대한 평활식으로 각 시기 t 에서 시계열의 추정된 수준을 얻을 수 있습니다.
- 평활식에서 ℓ_t 를 $\hat{y}_{t+1|t}$ 로 ℓ_{t-1} 을 $\hat{y}_{t|t-1}$ 로 바꾸면, 단순 지수평활의 가중 평균 형태를 얻을 수 있을 것입니다.
- 단순 지수 평활의 성분 형태는 특별히 유용하진 않습니다만, 다른 성분을 추가할 때 가장 쉽게 사용할 수 있는 형태일 것입니다.

6.1 단순 지수 평활

- 평평한 예측값
 - 단순 지수 평활은 “평평(flat)”한 예측 함수를 갖습니다:
$$\hat{y}_{T+h|T} = \hat{y}_{T+1|T} = \ell_T, \quad h = 2, 3, \dots$$
- 즉, 모든 예측값이 마지막 수준 성분과 같은 값을 갖습니다. 이러한 예측은 시계열에 추세나 계절 성분이 없을 때 사용할 수 있다는 사실을 기억합시다.
- 최적화

$$\text{SSE} = \sum_{t=1}^T (y_t - \hat{y}_{t|t-1})^2 = \sum_{t=1}^T e_t^2.$$

6.2 추세 기법

- 홀트의 선형 추세 기법

- Holt (1957) 에서는 추세가 있는 데이터를 예측할 수 있도록 단순 지수 평활을 확장했습니다. 이 방법은 예측식과 두 개의 평활식 (수준에 관한 것, 추세에 관한 것)을 포함합니다.

Forecast equation

$$\hat{y}_{t+h|t} = \ell_t + hb_t$$

Level equation

$$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1})$$

Trend equation

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1},$$

- 여기에서 ℓ_t 는 시간 t 에서 시계열의 수준 추정 값, b_t 는 시간 t 에서의 시계열의 추세(기울기) 추정 값, $0 \leq \alpha \leq 1$ 은 수준에 대한 매개변수, $0 \leq \beta^* \leq 1$ 은 추세에 대한 매개변수를 나타냅니다.
- 단순 지수평활처럼, 수준식은 ℓ_t 이 관측 y_t 의 가중 평균이라는 것과, 여기에서 $\ell_{t-1} + b_{t-1}$ 로 주어지는 시간 t 에 대한 한 단계 앞 학습 예측이라는 것을 나타냅니다. 추세식은 b_t 가 추세의 이전 추정 값인 $\ell_t - \ell_{t-1}$ 와 b_{t-1} 에 기초한, 시간 t 에서의 추정된 추세의 이동 평균이라는 것을 나타냅니다.

6.2 추세 기법

• 감쇠 추세 기법

- 홀트 선형 기법으로 얻은 예측값은 미래에도 계속 일정한 추세를 나타냅니다. 이러한 기법은 과도하게 예측하는 경향이 있고, 예측 범위(forecast horizon)이 늘어날 수록 더더욱 그렇습니다.

Forecast equation $\hat{y}_{t+h|t} = \ell_t + (\phi + \phi^2 + \cdots + \phi^h)b_t$

Level equation $\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$

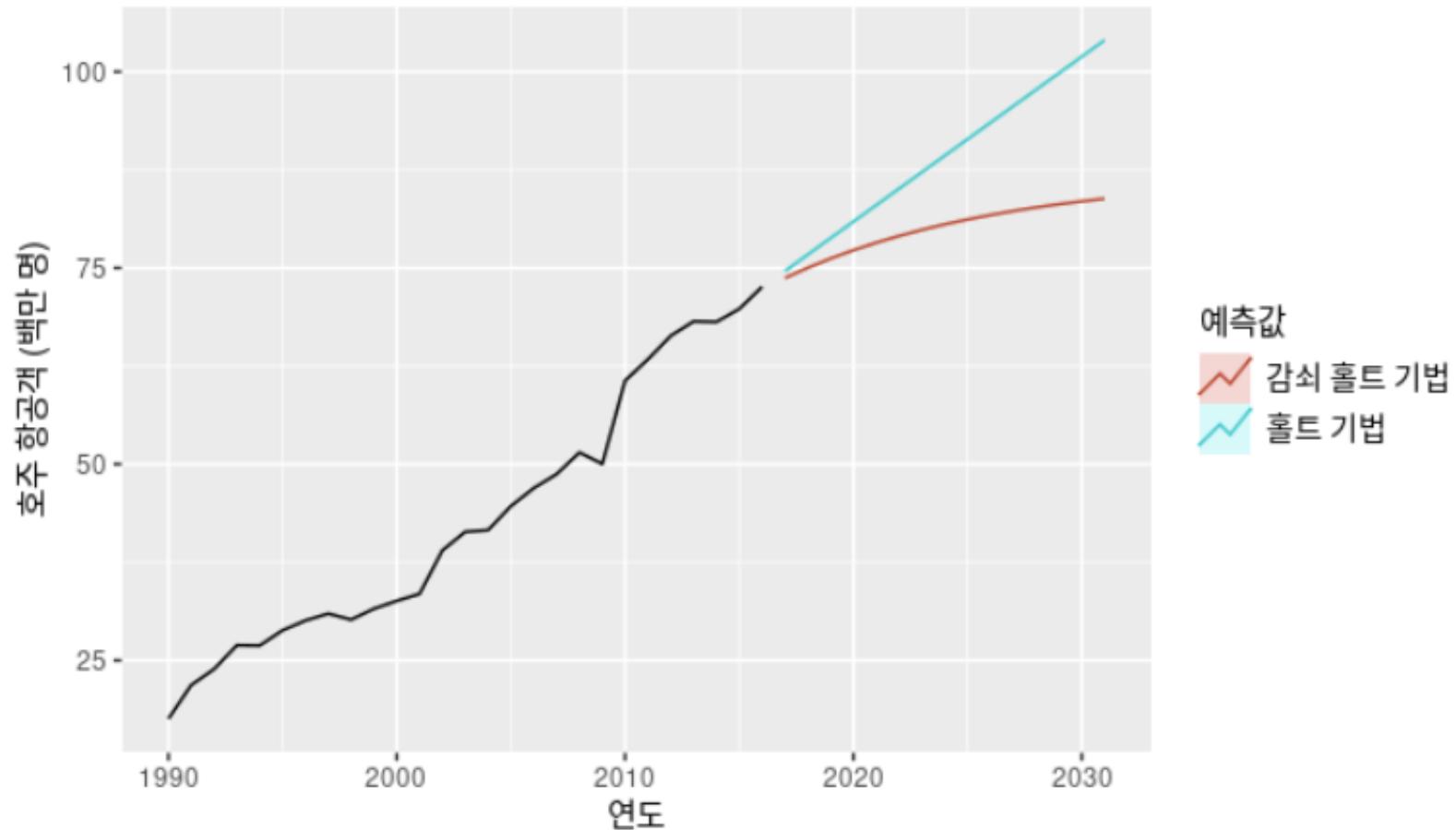
Trend equation $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$.

- 이러한 사실에 착안하여, Gardner & McKenzie ([1985](#))는 미래 어느 시점에 추세를 평평하게 감쇠시키는 한 가지 매개변수를 도입하였습니다.
- 만약에 $\phi = 1$ 이면, 이 기법은 홀트의 선형 기법과 완전히 같습니다. 0과 1 사이의 값에 대해 ϕ 는 추세를 감쇠시켜 미래 어떤 시점에 추세가 상수가 되도록 합니다. 사실 어떠한 $0 \leq \phi \leq 1$ 에 대해, $h \rightarrow \infty$ 일수록 예측치가 $\ell_t + \phi b_t / (1 - \phi)$ 로 수렴합니다. 이러한 효과에 의해 단기 예측값은 추세를 나타내고 장기 예측치는 상수가 됩니다.

보통 ϕ 의 최소값을 0.8로 잡고 최대값을 0.98로 제한합니다.

6.2 추세 기법

- 홀트기법으로 얻은 예측값



호주에서 등록된 항공사의 연간 전체 항공객 예측 (단위: 백만 명, 1990–2016).
감쇠 추세 기법에 대해, $\varphi=0.90$.

6.3 홀트-윈터스의 계절성 기법

- 홀트-윈터스(Holt-Winters) 계절성 기법
 - Holt (1957) 와 Winters (1960) 은 계절성을 잡아내기 위해 홀트 (Holt)의 기법을 확장하였습니다.

$$\hat{y}_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$$

$$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$$

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$$

$$s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m},$$

- 여기에서 k 는 $(h - 1)/m$ 의 정수 부분입니다. 이 값은 예측을 위해 계절성 지수를 추정한 값이 표본의 마지막 연도에서 유래하도록 합니다.
- 수준식은 계절성으로 조정된 관측 $(y_t - s_{t-m})$ 과 시간 t 에 대한 비-계절성 예측 $(\ell_{t-1} + b_{t-1})$ 을 나타냅니다.
- 홀트(Holt)의 선형 기법과 같습니다. 계절성식은 현재 계절성 지수 $(y_t - \ell_{t-1} - b_{t-1})$ 와 이전 연도 같은 계절(즉, m 시점 이전)의 계절성 지표 사이의 가중 평균을 나타냅니다.

6.3 홀트-윈터스의 계절성 기법

- 홀트-윈터스(Holt-Winters) 계절성 기법
 - 계절성 식은 종종 다음과 같이 나타냅니다.

$$s_t = \gamma^*(y_t - \ell_t) + (1 - \gamma^*)s_{t-m}.$$

- 평활식에서 ℓ_t 를 위의 성분 형태의 수준식에 대입하면, 다음과 같은 식을 얻습니다.

$$s_t = \gamma^*(1 - \alpha)(y_t - \ell_{t-1} - b_{t-1}) + [1 - \gamma^*(1 - \alpha)]s_{t-m},$$

- 위의 식은 여기에서 $\gamma = \gamma^*(1 - \alpha)$ 으로 명시한 계절성에 대한 평활식과 같습니다. 보통의 매개변수 제한조건은 $0 \leq \gamma^* \leq 1$ 인데, $0 \leq \gamma \leq 1 - \alpha$ 로 다시 쓸 수 있습니다.

6.3 홀트-윈터스의 계절성 기법

- 홀트-윈터스(Holt-Winters) 계절성 기법
 - 홀트-윈터스의 곱셈 기법

$$\hat{y}_{t+h|t} = (\ell_t + hb_t)s_{t+h-m(k+1)}$$

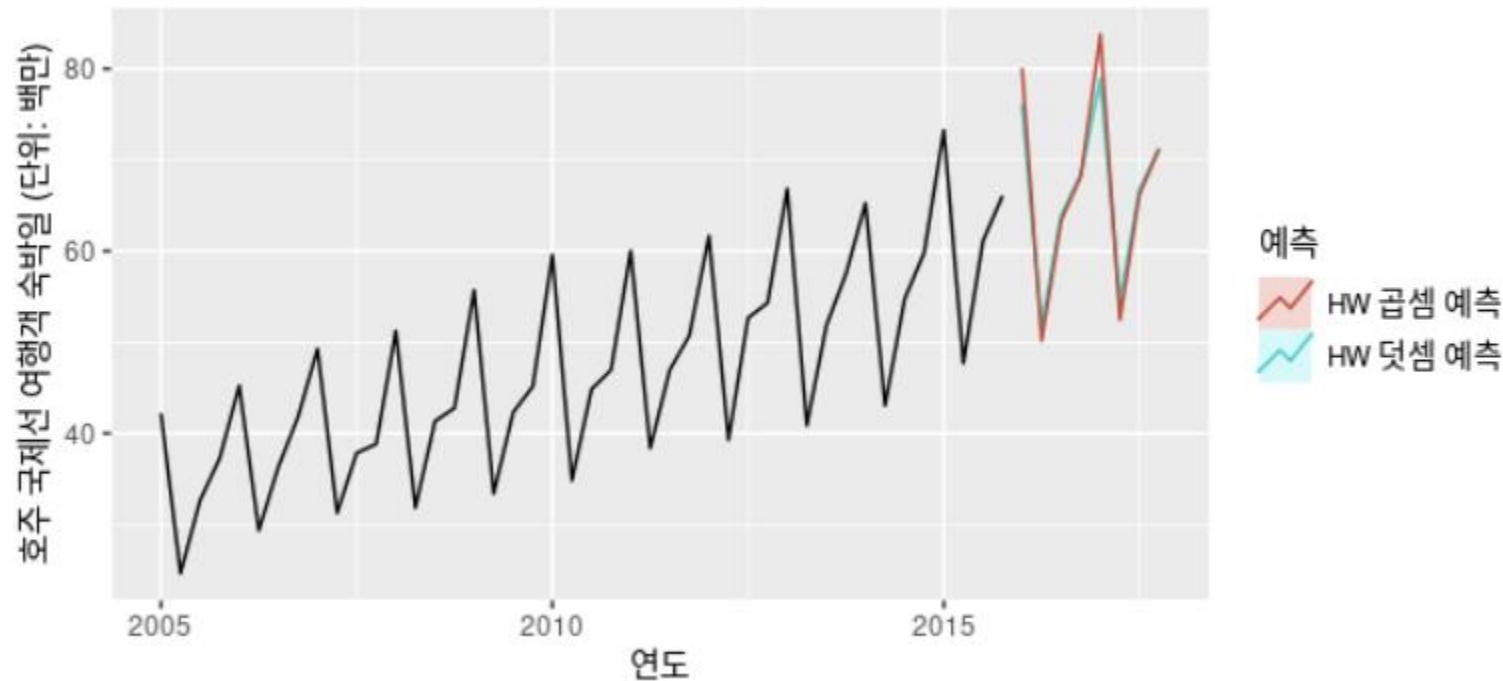
$$\ell_t = \alpha \frac{y_t}{s_{t-m}} + (1 - \alpha)(\ell_{t-1} + b_{t-1})$$

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$$

$$s_t = \gamma \frac{y_t}{(\ell_{t-1} + b_{t-1})} + (1 - \gamma)s_{t-m}$$

6.3 홀트-원터스의 계절성 기법

- 홀트-원터스 기법을 이용하여 호주를 방문하는 국제선 여행객을 예측



6.3 홀트-윈터스의 계절성 기법

- 홀트-윈터스의 감쇠 기법
 - 홀트-윈터스(Holt-Winters)의 덧셈과 곱셈 기법 두 경우 모두 감쇠 효과를 추가할 수 있습니다.
 - 계절성 데이터에 대해 정확하고 안정적인 예측치를 내는 한 가지 기법은 다음과 같이 홀트-윈터스(Holt-Winters)에 감쇠 추세 (damped trend)와 곱셈 계절성(multiplicative seasonality)을 고려한 것입니다:

$$\hat{y}_{t+h|t} = [\ell_t + (\phi + \phi^2 + \cdots + \phi^h)b_t] s_{t+h-m(k+1)}.$$

$$\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$$

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$$

$$s_t = \gamma \frac{y_t}{(\ell_{t-1} + \phi b_{t-1})} + (1 - \gamma)s_{t-m}.$$

6.4 지수 평활 기법 분류 체계

- 추세와 계절적인 성분의 조합을 고려한 지수 평활 기법을 2가지로 분류

추세 성분	계절 성분		
	N (없음)	A (덧셈)	M (곱셈)
N (없음)	(N,N)	(N,A)	(N,M)
A (덧셈)	(A,N)	(A,A)	(A,M)
A_d (덧셈 감쇠)	(A_d ,N)	(A_d ,A)	(A_d ,M)

Short hand	Method
(N,N)	단순 지수 평활
(A,N)	홀트의 선형 기법
(A_d ,N)	덧셈 감쇠 추세 기법
(A,A)	덧셈 홀트-윈터스 기법
(A,M)	곱셈 홀트-윈터스 기법
(A_d ,M)	홀트-윈터스 감쇠 기법

6.5 지수 평활에 대한 혁신 상태 공간 모델

- 통계적 모델
 - 지수 평활 기법은 점 예측값을 내는 알고리즘이며, 통계모델은 같은 점 예측값을 내는 동시에 예측 구간도 생성
 - 통계 모델이란 전체 예측분포를 만들어줄 수 있는 무작위적 데이터 생성 과정
 - 각 모델은 관측된 데이터를 묘사하는 측정식(measurement equation)과, 아직 관측되지 않은 성분이나 상태(수준, 추세, 계절성)가 시간에 따라 어떻게 변하는지 기술하는 몇 가지 상태식(state equation)으로 구성됩니다.
 - 이러한 이유에서 상태 공간 모델(state space models)이라고 부릅니다.

6.5 지수 평활에 대한 혁신 상태 공간 모델

- 통계적 모델
 - 각 기법마다 두 가지 모델이 존재합니다. 하나는 덧셈 오차를, 다른 하나는 곱셈 오차를 이용합니다. 같은 평활 매개변수 값을 사용했다면, 모델이 낸 얻은 점 예측은 같습니다. 하지만, 모델은 다른 예측 구간을 생성할 것입니다.
 - 덧셈 오차와 곱셈 오차를 사용하는 모델을 구분하기 위해(그리고 또한 기법에서 모델을 구분하기 위해) 지수 평활 기법 분류 체계표의 분류에 세 번째 문자 하나를 더 추가합니다. (오차Error, 추세Trend, 계절성Seasonal)에 각 상태 공간 모델은 ETS(\cdot, \cdot, \cdot)로 나타냅니다.
 - ETS: ExponenTial Smoothing
 - 각 성분에 대한 가능 상태: 오차 = {A,M}, 추세 = {N,A,Ad} 그리고 계절성 = {N,A,M}.

6.5 지수 평활에 대한 혁신 상태 공간 모델

- 통계적 모델
 - 지수 평활 기법은 점 예측값을 내는 알고리즘이며, 통계모델은 같은 점 예측값을 내는 동시에 예측 구간도 생성
 - 통계 모델이란 전체 예측분포를 만들어줄 수 있는 무작위적 데이터 생성 과정
 - 각 모델은 관측된 데이터를 묘사하는 측정식(measurement equation)과, 아직 관측되지 않은 성분이나 상태(수준, 추세, 계절성)가 시간에 따라 어떻게 변하는지 기술하는 몇 가지 상태식(state equation)으로 구성됩니다.
 - 이러한 이유에서 상태 공간 모델(state space models)이라고 부릅니다.

6.5 지수 평활에 대한 혁신 상태 공간 모델

- ETS(A,N,N): 덧셈 오차를 이용하는 단순 지수평활

$$\text{Forecast equation} \quad \hat{y}_{t+1|t} = \ell_t$$

$$\text{Smoothing equation} \quad \ell_t = \alpha y_t + (1 - \alpha) \ell_{t-1},$$

- 수준에 대한 평활식을 다시 정리하면, “오차 보정(error correction)” 식을 얻습니다:

$$\begin{aligned}\ell_t &= \ell_{t-1} + \alpha(y_t - \ell_{t-1}) \\ &= \ell_{t-1} + \alpha e_t\end{aligned}$$

- 여기에서 $t = 1, \dots, T$ 에 $e_t = y_t - \ell_{t-1} = y_t - \hat{y}_{t|t-1}$ 는 시간 t 에서의 잔차(residual)입니다.
 - 훈련 데이터 오차는 $t = 1, \dots, T$ 에 대한 평활 과정에 걸쳐 추정된 수준의 조정으로 이어집니다.
 - 예를 들면, 시간 t 에서 오차가 음수이면, $y_t < \hat{y}_{t|t-1}$ 이고 따라서 시간 $t - 1$ 에서 수준은 과도하게 추정됩니다. 그러면 새로운 수준 ℓ_t 은 하향 조정된 이전 수준 ℓ_{t-1} 이 됩니다.
 - α 가 1에 가까울 수록, 수준 추정이 더 고르지 않게 됩니다(큰 조정이 일어납니다). α 가 작을 수록, 수준이 더 고르게 됩니다(작은 조정이 일어납니다).

6.5 지수 평활에 대한 혁신 상태 공간 모델

- ETS(A,N,N): 덧셈 오차를 이용하는 단순 지수평활

$$\begin{aligned}\ell_t &= \ell_{t-1} + \alpha(y_t - \ell_{t-1}) \\ &= \ell_{t-1} + \alpha e_t\end{aligned}$$

- 각 관측값이 이전 수준에 오차를 더한 것과 같게 두기 위해 다음과 같이 $y_t = \ell_{t-1} + e_t$ 이렇게 쓸 수도 있습니다.
- 이것을 혁신 상태 공간 모델(innovation state space model)로 만들기 위해, e_t 대한 확률 분포를 식으로 구체적으로 적는 작업이 필요합니다.
- 덧셈 오차를 이용하는 모델에 대해, 잔차(한 단계 학습 오차) e_t 가 평균이 0이면서 분산이 σ^2 인 정규 분포를 따르는 백색잡음(white noise)이라고 가정합니다.
- 이렇게 가정한 것을 다음과 같이 $e_t = \varepsilon_t \sim NID(0, \sigma^2)$ 이렇게 간단하게 씁니다.
- 여기에서 NID는 “정규적으로 그리고 독립적으로 분포된(normaly and independently distributed)”이라는 말을 줄여 쓴 것입니다.

6.5 지수 평활에 대한 혁신 상태 공간 모델

- ETS(A,N,N): 덧셈 오차를 이용하는 단순 지수평활
 - 그러면 모델의 식들을 다음과 같이 쓸 수 있습니다.

$$y_t = \ell_{t-1} + \varepsilon_t$$
$$\ell_t = \ell_{t-1} + \alpha \varepsilon_t.$$

- y_t 식을 측정(또는 관측) 방정식으로, ℓ_t 식을 상태(또는 전이) 방정식으로 부릅시다.
- 오차의 통계적인 분포를 함께 이용한 이러한 두 식이 전체적으로 명확한 통계 모델을 이룹니다. 특별히, 이러한 식이 단순 지수 평활을 이루는 혁신 상태 공간 모델(innovations state space model)이 됩니다.
 - “혁신(innovations)”이라는 단어는 이러한 종류의 설명에서 모든 식이 같은 무작위 오차 과정 ε_t 을 사용한다는 사실에서 유래합니다. 같은 이유에서, 이렇게 식을 세우는 과정을 “오차의 단일 원천(single source of error)”으로 부르기도 합니다. 오차가 생기는 다른 여러 이유가 있습니다 (여기에서 다루지는 않았습니다).
- 측정 방정식은 관측값과 아직 관측되지 않은 상태와의 관계를 보여줍니다. 이 경우에, 관측값 y_t 는 수준 ℓ_{t-1} 과, y_t 의 예측 가능한 부분과, 오차 ε_t , y_t 의 예측 가능하지 않은 부분의 선형 함수입니다. 다른 혁신 상태 공간 모델의 경우에는 이러한 관계가 비선형적일 수도 있습니다.

6.5 지수 평활에 대한 혁신 상태 공간 모델

- ETS(M,N,N): 곱셈 오차를 이용하는 단순 지수평활
 - 같은 방식으로, 한 단계 앞 학습 오차를 상대적인 오차로 써서 곱셈 오차를 이용하는 모델을 세울 수 있습니다:

$$\varepsilon_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}}$$

- 여기에서 $\varepsilon_t \sim NID(0, \sigma^2)$ 입니다. $\hat{y}_{t|t-1} = \ell_{t-1}$ 을 대입하면, $y_t = \ell_{t-1} + \ell_{t-1}\varepsilon_t$ 와 $e_t = y_t - \hat{y}_{t|t-1} = \ell_{t-1}\varepsilon_t$ 을 얻습니다.
- 그러면 상태 공간 모델의 곱셈 형태를 다음과 같이 쓸 수 있습니다.

$$y_t = \ell_{t-1}(1 + \varepsilon_t)$$
$$\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t).$$

6.5 지수 평활에 대한 혁신 상태 공간 모델

- ETS(A,A,N): 덧셈 오차를 이용한 홀트의 선형 기법

- 이 모델에서, 한 단계 앞 학습 오차가 $\varepsilon_t = y_t - \ell_{t-1} - b_{t-1} \sim NID(0, \sigma^2)$ 이렇게 주어진다고 가정합니다.
 - 이것을 홀트(Holt)의 선형 기법에 대한 오차 보정식에 대입하면 다음과 같은 식을 얻습니다.

$$\begin{aligned}y_t &= \ell_{t-1} + b_{t-1} + \varepsilon_t \\ \ell_t &= \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t \\ b_t &= b_{t-1} + \beta \varepsilon_t,\end{aligned}$$

- 간결하게 쓰기 위해, 여기에서 $\beta = \alpha \beta^*$ 로 두었습니다

6.5 지수 평활에 대한 혁신 상태 공간 모델

- ETS(M,A,N): 곱셈 오차를 이용하는 홀트의 선형 기법
 - 한 단계 앞 학습 오차를 상대적인 오차로 다음과 같이 적고

$$\varepsilon_t = \frac{y_t - (\ell_{t-1} + b_{t-1})}{(\ell_{t-1} + b_{t-1})}$$

- 위에서 사용한 것과 비슷한 접근 방식을 따르면, 곱셈 오차를 이용하는 홀트의 선형 기법을 이루는 혁신 상태 공간 모델을 다음과 같이 구체적으로 적을 수 있습니다.

$$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$$

$$\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$$

$$b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$$

- 다시 한번 여기에서 $\beta = \alpha\beta^*$ 이고 $\varepsilon_t \sim NID(0, \sigma^2)$ 입니다.

6.5 지수 평활에 대한 혁신 상태 공간 모델

- 다른 덧셈 ETS 모델
 - 비슷한 방식으로, 지수평활(exponential smoothing) 모델 각각에 대한 혁신 상태 공간 모델(innovation state space model)을 쓸 수 있습니다.

덧셈 오차 모델

추세	계절성		
	N	A	M
N	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = \ell_{t-1} s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / \ell_{t-1}$
A	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1}) s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = b_{t-1} + \beta \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + b_{t-1})$
A _d	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1}) s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = \phi b_{t-1} + \beta \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + \phi b_{t-1})$

6.5 지수 평활에 대한 혁신 상태 공간 모델

- 다른 곱셈 ETS 모델
 - 비슷한 방식으로, 지수평활(exponential smoothing) 모델 각각에 대한 혁신 상태 공간 모델(innovation state space model)을 쓸 수 있습니다.

곱셈 오차 모델

추세	계절성		
	N	A	M
N	$y_t = \ell_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$	$y_t = (\ell_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A	$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A _d	$y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

6.6 추정과 모델 선택

- ETS 모델들 추정하기
 - 제곱 오차의 합을 최소화하여 매개변수를 추정하는 방법의 한 가지 대안은 “가능도(또는 우도)likelihood”를 최대화하는 것입니다. 가능도(likelihood)는 특정한 모델에서 일어나는 데이터의 확률입니다.
 - 그래서 큰 가능도는 좋은 모델과 관련 있습니다. 덧셈 오차 모델의 경우에, (오차가 정규 분포를 따른다고 가정하고) 가능도를 최대화하면 제곱 오차의 합을 최소화하는 것과 결과가 같습니다. 하지만, 곱셈 오차 모델의 경우에는 서로 다른 결과를 낼 것입니다.
 - 가능도를 최대화하여 평활 매개변수 $\alpha, \beta, \gamma, \phi$ 과 초기 상태 $\ell_0, b_0, s_0, s_{-1}, \dots, s_{-m+1}$ 을 구할 것입니다.
 - 평활 매개변수가 가질 수 있는 값의 범위는 제한되어 있습니다. 전통적으로는 관련 식이 가중평균으로 해석될 수 있도록 매개변수가 0과 1사이 범위로 제한됩니다.
 - 즉, $0 < \alpha, \beta^*, \gamma^*, \phi < 1$ 입니다.
 - 상태 공간 모델의 경우에, $\beta = \alpha\beta^*$ 와 $\gamma = (1 - \alpha)\gamma^*$ 로 두었습니다.
 - 그래서 전통적인 제한조건이, $0 < \alpha < 1, 0 < \beta < \alpha, 0 < \gamma < 1 - \alpha$ 로 바뀝니다. 실제적인 상황에서는 수치적인 어려움을 막기 위해 보통은 감쇠 매개변수 ϕ 에도 제한을 둡니다.
 - R프로그램에서는 $0.8 < \phi < 0.98$ 가 되도록 제한합니다.

6.6 추정과 모델 선택

- ETS 모델들 추정하기

- 상태 공간 모델의 수학적인 특성을 고려하는 것은 매개변수를 해석하는 또 다른 방법입니다.
- 현재 예측값에 지속적으로 영향을 미치는 먼 과거의 관측값을 방지하도록 매개변수에 제한 조건이 걸립니다. 이로부터 매개변수에 대한 몇 가지 허용성(*admissibility*) 제한 조건이 나옵니다.
- 이러한 제한 조건은 (항상 그렇지는 않지만) 보통은 전통적인 제한조건의 범위보다는 덜 제한적입니다.
 - 예를 들면, ETS(A,N,N) 모델의 경우에는, 흔히 사용하는 매개변수 범위는 $0 < \alpha < 1$ 이지만, 허용 범위는 $0 < \alpha < 2$ 입니다.
 - ETS(A,A,N) 모델의 경우에는, 전통적인 매개변수 범위는 $0 < \alpha < 1$ 와 $0 < \beta < \alpha$ 이지만, 허용 범위는 $0 < \alpha < 2$ 와 $0 < \beta < 4 - 2\alpha$ 입니다.

6.6 추정과 모델 선택

- 모델 선택
 - 모델을 선택할 때 정보 기준을 사용할 수 있다는 것은 ETS 통계 체계의 큰 장점입니다.
 - ETS 모델의 경우에 아카이케(Akaike)의 정보 기준(AIC)은 다음과 같이 정의됩니다.

$$AIC = -2 \log(L) + 2k,$$

- 여기에서 L 은 모델의 가능도(likelihood)이고, k 는 잔차 분산을 포함하여 추정되는 매개변수와 초기 상태의 전체 개수입니다.
- 작은 표본에 대해 보정된 AIC는(AIC_c) 다음과 같이 정의됩니다.

$$AIC_c = AIC + \frac{2k(k+1)}{T-k-1},$$

- 그리고 베이지안(Bayesian) 정보 기준(BIC)은

$$BIC = AIC + k[\log(T) - 2].$$

(오차Error, 추세Trend, 계절성Seasonal)의 세 가지 성분의 조합에서 수치적으로 어려운 부분이 생길 수 있습니다. 구체적으로, 이러한 불안정성을 일으킬 수 있는 모델은 ETS(A,N,M), ETS(A,A,M), ETS(A,A_d,M)입니다. 상태 식에서 0에 가까운 값으로 나눌 수도 있기 때문입니다. 이런 특정 조합은 모델을 선택할 때 보통은 고려하지 않겠습니다.

Thank you

Q & A

