10.1 Quarterly single-family housing starts. The quarterly numbers of single-family housing starts (in thousands of dwellings) in the United States from 2004 through 2008 are recorded in the next table (p. 535).

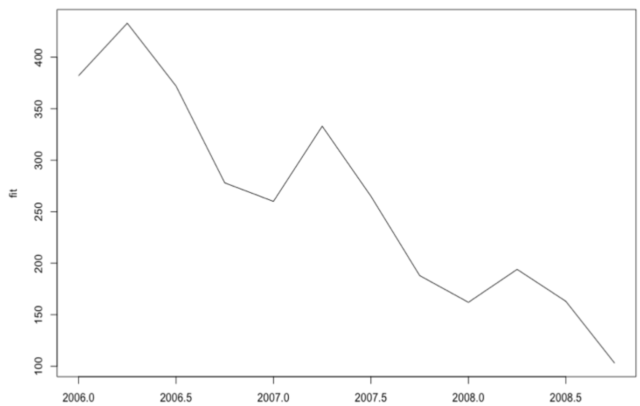
QTRHOUSE

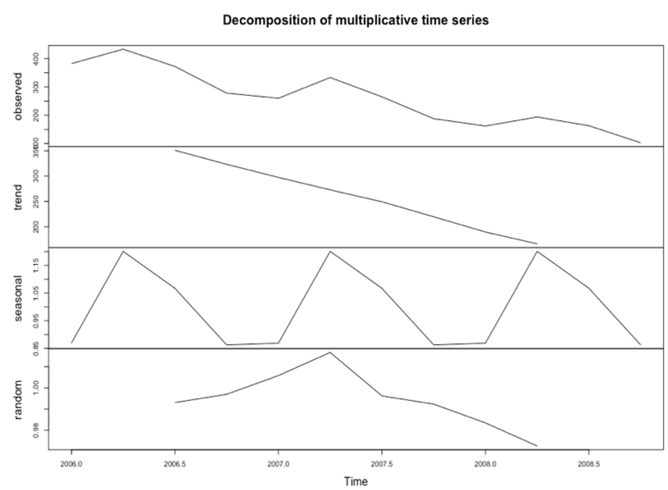
YEAR	QUARTER	HOUSING STARTS
2006	1	382
	2	433
	3	372
	4	278
2007	1	260
	2	333
	3	265
	4	188
2008	1	162
	2	194
	3	163
	4	103

Source: U.S. Bureau of the Census, *Statistical Abstract of*

(a) Plot the quarterly time series. Can you detect a long-term trend? Can you detect any seasonal variation?

```
> input<-c(382,433,372,278,260,333,265,188,162,194,163,103)
> fit<-ts(input,frequency=4,start=c(2006,1))
> plot(fit)
> decompose(fit,type="mult")
$x
     Qtr1 Qtr2 Qtr3 Qtr4
2006  382  433  372  278
2007  260  333  265  188
2008  162  194  163  103

$seasonal
          Qtr1      Qtr2      Qtr3      Qtr4
2006 0.8692790 1.2008116 1.0672551 0.8626543
2007 0.8692790 1.2008116 1.0672551 0.8626543
2008 0.8692790 1.2008116 1.0672551 0.8626543

$trend
        Qtr1    Qtr2    Qtr3    Qtr4
2006      NA      NA 351.000 323.250
2007 297.375 272.750 249.250 219.625
2008 189.500 166.125      NA      NA

$random
          Qtr1      Qtr2      Qtr3      Qtr4
2006        NA        NA 0.9930419 0.9969411
2007 1.0057955 1.0167276 0.9961907 0.9922916
2008 0.9834371 0.9725050        NA        NA

$figure
[1] 0.8692790 1.2008116 1.0672551 0.8626543

$type
[1] "multiplicative"

attr(,"class")
[1] "decomposed.ts"
> decomposedRes<-decompose(fit,type="mult")
> plot(decomposedRes)
```





Ans: a long-term downward trend and a seasonal trend peaking at Qtr 2 are observed.

10.11 Graphing calculator sales. The table below presents the quarterly sales index for one brand of graphing calculator at a campus bookstore. The quarters are based on an academic year, so the first quarter represents fall; the second, winter; the third, spring; and the fourth, summer.

Define the time variable as $t = 1$ for the first quarter of 2005, $t = 2$ for the second quarter of 2005, etc. Consider the following seasonal dummy variables:

$$Q_1 = \begin{cases} 1 & \text{if Quarter 1} \\ 0 & \text{otherwise} \end{cases}$$

$$Q_2 = \begin{cases} 1 & \text{if Quarter 2} \\ 0 & \text{otherwise} \end{cases}$$

$$Q_3 = \begin{cases} 1 & \text{if Quarter 3} \\ 0 & \text{otherwise} \end{cases}$$

GRAPHICAL

YEAR	FIRST QUARTER	SECOND QUARTER	THIRD QUARTER	FOURTH QUARTER
2005	438	398	252	160
2006	464	429	376	216
2007	523	496	425	318
2008	593	576	456	398
2009	636	640	526	498

(a) Write a regression model for $E(Y_t)$ as a function of t, Q_1, Q_2, and Q_3.

Ans: $\boldsymbol{E(y_t) = \beta_0 + \beta_1 t + \beta_2 Q_1 + \beta_3 Q_2 + \beta_4 Q_3}$

(b) Find and interpret the least squares estimates and evaluate the usefulness of the model.

$H_0\colon \beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0;\ H_a\colon$ one of the $\beta_i' s$ is nonzero

```
> y<-c(438,398,252,160,464,429,376,216,523,496,425,318,593,576,456,398,6
36,640,526,498)
> q1<-c(1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0)
> q2<-c(0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0)
> q3<-c(0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0,0,0,1,0)
> t<-c(1:20)
> calculatorLM<-lm(y~t+q1+q2+q3)
> summary(calculatorLM)

Call:
lm(formula = y ~ t + q1 + q2 + q3)

Residuals:
   Min     1Q Median     3Q    Max 
-35.95 -14.09  -2.30  14.96  47.90 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  119.850     16.950   7.071 3.80e-06 ***
t             16.513      1.028  16.067 7.33e-11 ***
q1           262.338     16.730  15.681 1.04e-10 ***
q2           222.825     16.571  13.446 8.99e-10 ***
q3           105.512     16.476   6.404 1.19e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 26 on 15 degrees of freedom
Multiple R-squared:  0.9692,	Adjusted R-squared:  0.9609 
F-statistic: 117.8 on 4 and 15 DF,  p-value: 3.86e-11
```

Ans: $\boldsymbol{\hat{y}_t = 119.850 + 16.513t + 262.338Q_1 + 222.825Q_2 + 105.512Q_3,}$

$\boldsymbol{F = 117.8, p\text{- value} = 3.86 \times 10^{-11}}$ **(reject** $\boldsymbol{H_0}$**)**

(c) Which of the assumptions about the random error component is in doubt when a regression model is fit to time series data?

Ans: independent error

(d) Find the forecasts and the 95% prediction intervals for the 2010 quarterly sales. Interpret the result.

```
> predict(calculatorLM,newdata=data.frame(t=21,q1=1,q2=0,q3=0),interval=
"prediction")
     fit      lwr      upr
1 728.95 662.7977 795.1023
> predict(calculatorLM,newdata=data.frame(t=22,q1=0,q2=1,q3=0),interval=
"prediction")
     fit      lwr      upr
1 705.95 639.7977 772.1023
> predict(calculatorLM,newdata=data.frame(t=23,q1=0,q2=0,q3=1),interval=
"prediction")
     fit      lwr      upr
1 605.15 538.9977 671.3023
> predict(calculatorLM,newdata=data.frame(t=24,q1=0,q2=0,q3=0),interval=
"prediction")
     fit      lwr      upr
1 516.15 449.9977 582.3023
```

Ans:

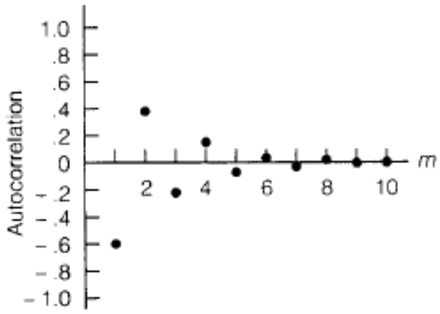
$\boldsymbol{Q_1: 728.95, (662.7997, 795.1023)}$

$\boldsymbol{Q_2: 705.95, (639.7997, 772.1023)}$

$\boldsymbol{Q_3: 605.15, (538.9977, 671.3023)}$

$\boldsymbol{Q_4: 516.115, (449.9977, 582.3023)}$

10.18 Identifying the autoregressive model. Consider the autoregression pattern shown in the figure. Write a first-order autoregressive model that exhibits this pattern.



Ans: $\boldsymbol{R_t = -0.6R_{t-1} + \varepsilon_t}$

10.20 Modeling stock price. Suppose you are interested in buying stock in the Pepsi Company (PepsiCo). Your broker has advised you that your best strategy is to sell the stock at the first substantial jump in price. Hence, you are interested in a short-term investment. Before buying, you would like to model the closing price of PepsiCo, y_t, over time (in days), t.

(a) Write a first-order model for the deterministic portion of the model, $E(y_t)$.

Ans: $\boldsymbol{E(y_t) = \beta_0 + \beta_1 t}$

(b) If a plot of the daily closing prices for the past months reveals a quadratic trend, write a plausible model for $E(y_t)$.

Ans: $\boldsymbol{E(y_t) = \beta_0 + \beta_1 t + \beta_2 t^2}$

(c) Since the closing price of PepsiCo on day $(t + 1)$ is very highly correlated with the closing price on day t, your broker suggests that the random error components of the model are not white noise. Given this information, postulate a model for the error term, R_t.

Ans: $\boldsymbol{R_t = \phi R_{t-1} + \varepsilon_t}$

10.22 Overbooking airline flights. Airlines sometimes over flights of "no-show" passengers (ie., passengers who have purchased a ticket but fail to board the flight). An airline supervisor wishes to be able to predict, for a flight from Miami to New York, the monthly accumulation of no-show passengers during the upcoming year, using data from the past 3 years. Let y_t = Number of no-shows during month t.

(a) Using dummy variables, propose a model for $E(y_t)$ that will take into account the seasonal (fall, winter, spring, summer) variation that may be present in the data.

Ans: $\boldsymbol{E(y_t) = \beta_0 + \beta_1 S_1 + \beta_2 S_2 + \beta_3 S_3 + \beta_4 t}$

$$\boldsymbol{S_1 = \begin{cases} 1 & \text{if season is spring (II)} \\ 0 & \text{otherwise} \end{cases}, S_2 = \begin{cases} 1 & \text{if season is summer (III)} \\ 0 & \text{otherwise} \end{cases}}$$

$$\boldsymbol{S_3 = \begin{cases} 1 & \text{if season is fall (III)} \\ 0 & \text{otherwise} \end{cases}}$$

(b) Postulate a model for the error term R_t.

Ans: $\boldsymbol{R_t = \phi R_{t-1} + \varepsilon_t}$

(c) Write the full time series model for y_t (include random error terms).

Ans: $\boldsymbol{y_t = \beta_0 + \beta_1 S_1 + \beta_2 S_2 + \beta_3 S_3 + \beta_4 t + \phi R_{t-1} + \varepsilon_t}$

(d) Suppose the airline supervisor believes that the seasonal variation in the data is not constant from year to year, in other words, that there exists interaction between time and season. Rewrite the full model with the interaction terms added.

Ans: $\boldsymbol{y_t = \beta_0 + \beta_1 S_1 + \beta_2 S_2 + \beta_3 S_3 + \beta_4 t + \beta_5 S_1 t + \beta_6 S_2 t + \beta_7 S_3 t + \phi R_{t-1} + \varepsilon_t}$

10.25 Quarterly GDP values. The gross domestic product (GDP) is a measure of total U.S. output and is, therefore, an important indicator of the U.S. economy. The quarterly GDP values (in billion of dollars) from 2004 to 2008 are given in the next table (p. 558). Let y_t be the GDP in quarter t, $t = 1, 2, 3, \ldots, 20$.

GDP

YEAR	QUARTER	GDP
2004	1	11,406
	2	11,610
	3	11,779
	4	11,949
2005	1	12,155
	2	12,298
	3	12,538
	4	12,695
2006	1	12,960
	2	13,134
	3	13,250
	4	13,370
2007	1	13,511
	2	13,738
	3	13,951
	4	14,031
2008	1	14,151
	2	14,295
	3	14,413
	4	14,200

Source: U.S. Department of Commerce, Bureau of Economic Analysis, 2009; www.bea.gov.

(a) Hypothesize a time series model for quarterly GDP that includes a straight-line long-term trend and autocorrelated residuals.

Ans: $\boldsymbol{y_t = \beta_0 + \beta_1 t + \phi R_{t-1} + \varepsilon_t}$

(b) The SAS printout for the time series model $y_t = \beta_0 + \beta_1 t + \phi R_{t-1} + \varepsilon_t$ is shown at the bottom of p. 558. Write the least squares prediction equation.

SAS Output for Exercise 10.25

```
                      The AUTOREG Procedure

                   Dependent Variable    GDP

                 Ordinary Least Squares Estimates

  SSE              297252.367    DFE                        18
  MSE                   16514    Root MSE            128.50689
  SBC              254.881096    AIC                252.889631
  MAE              85.8141203    AICC               253.595514
  MAPE             0.65182465    Regress R-Square       0.9834
  Durbin-Watson        0.6170    Total R-Square         0.9834

                                     Standard                Approx
 Variable      DF      Estimate         Error    t Value    Pr > |t|

 Intercept      1         11363       59.6955     190.35      <.0001
 T              1      162.7266        4.9833      32.65      <.0001

                  Estimates of Autocorrelations

Lag   Covariance   Correlation   -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

  0      14862.6      1.000000   |                    |********************|
  1       5563.1      0.374304   |                    |*******             |

                  Preliminary MSE      12780.3

             Estimates of Autoregressive Parameters

                                     Standard
             Lag      Coefficient       Error      t Value

               1        -0.374304    0.224905        -1.66

                     Yule-Walker Estimates

  SSE               225352.96    DFE                        17
  MSE                   13256    Root MSE            115.13495
  SBC              252.489499    AIC                249.502303
  MAE              69.2335112    AICC               251.002303
  MAPE             0.52383884    Regress R-Square       0.9724
  Durbin-Watson        0.7900    Total R-Square         0.9874

                                     Standard                Approx
 Variable      DF      Estimate         Error    t Value    Pr > |t|

 Intercept      1         11374       79.4734     143.12      <.0001
 T              1      160.2342        6.5421      24.49      <.0001
```

Ans: $\boldsymbol{\hat{y}_t = 11{,}374 + 160.23t + 0.3743\hat{R}_{t-1}}$

(c) Interpret the estimates of the model parameters, β_0, β_1, and ϕ.

Starting at 11,374 billion dollars in Q1 2004, the quarterly GDP value is estimated to increase by 160.23 billon dollars per fiscal quarter with a correlated error of 0.3743 billion dollars.

(d) Interpret the values of R^2 and S.

98.74% of the variations can be explained by the time-series model and the model predictions will usually be accurate to within approximately ±2(115.13495) billon dollars.

Ans: $\boldsymbol{R^2 = 0.9874, s = 115.13495}$