CHAPTER 3 TRENDS（趋势）

3.1 Deterministic Versus Stochastic Trends

3.2 Estimation of a Constant Mean

3.3 Regression Methods

3.4 Reliability and Efficiency of Regression Estimates

3.5 Interpreting Regression Output

3.6 Residual Analysis

3.7 Summary

3.1 Deterministic Versus Stochastic Trends

stochastic trends e.g., random walk

A second simulation of exactly the same process might well show completely **different** "trends."

What is the different between the two concepts?

deterministic function $E(X_t) = 0$ for all t

deterministic trends $Y_t = \mu_t + X_t$

e.g., average monthly temperature series

the **reason** for the trend is **clear.** trend applies forever

3.2 Estimation of a Constant Mean

model:

$$Y_t = \mu + X_t$$

where $E(X_t) = 0$ for all t.

estimate of μ $$\bar{Y} = \frac{1}{n}\sum_{t=1}^{n} Y_t$$

$E(\bar{Y}) = \mu$ → unbiased estimate

the precision of $\bar{Y}$ as an estimate of μ?

3.2 Estimation of a Constant Mean

Suppose that $\{Y_t\}$, (or, equivalently, $\{X_t\}$) is a stationary time series with autocorrelation function ρ_k.

Exercise 2.17

$$Var(\bar{Y}) = \frac{\gamma_0}{n}\left[\sum_{k=-n+1}^{n-1}\left(1-\frac{|k|}{n}\right)\rho_k\right] = \frac{\gamma_0}{n}\left[1+2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\rho_k\right]$$

Case 1:

the series $\{X_t\}$ is just white noise, then $\rho_k = 0$ for $k > 0$

$$Var(\bar{Y}) = \frac{\gamma_0}{n}$$

3.2 Estimation of a Constant Mean

Case 2:

(stationary) moving average model $X_t = e_t - \frac{1}{2}e_{t-1}$

we find that $\rho_1 = -0.4$ and $\rho_k = 0$ for $k > 1$.

$$Var(\bar{Y}) = \frac{\gamma_0}{n}\left[1 + 2\left(1 - \frac{1}{n}\right)(-0.4)\right]$$

$$= \frac{\gamma_0}{n}\left[1 - 0.8\left(\frac{n-1}{n}\right)\right] \approx 0.2\frac{\gamma_0}{n}$$

Why?

Negative correlation at lag 1 has **improved** the estimation of the mean compared with the estimation obtained in the white noise situation.

3.2 Estimation of a Constant Mean

Case 3:

if $\rho_k \geq 0$ for all $k \geq 1$,

$$Var(\bar{Y}) = \frac{\gamma_0}{n}\left[1 + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\rho_k\right] \geq \frac{\gamma_0}{n}$$

The **positive** correlations make estimation of the mean ***more*** **difficult** than in the white noise case.

Why?

3.2 Estimation of a Constant Mean

For many stationary processes

$$\sum_{k=0}^{\infty} |\rho_k| < \infty \Rightarrow Var(\bar{Y}) \approx \frac{\gamma_0}{n}\left[\sum_{k=-\infty}^{\infty} \rho_k\right]$$

suppose that $\rho_k = \phi^{|k|}$ for all k

$$Var(\bar{Y}) \approx \frac{(1+\phi)}{(1-\phi)}\frac{\gamma_0}{n}$$

3.2 Estimation of a Constant Mean

For a nonstationary process (but with a constant mean), the precision of the sample mean as an estimate of μ can be strikingly different.

e.g., $\{X_t\}$ is a random walk process

$$
\begin{aligned}
Var(\bar{Y}) &= \frac{1}{n^2} Var\left[\sum_{i=1}^{n} Y_i\right] = \frac{1}{n^2} Var\left[\sum_{i=1}^{n}\sum_{j=1}^{i} e_j\right] \\
&= \frac{1}{n^2} Var(e_1 + 2e_2 + 3e_3 + \cdots + ne_n) = \frac{\sigma_e^2}{n^2} \sum_{k=1}^{n} k^2 \\
&= \sigma_e^2 (2n+1)\frac{(n+1)}{6n}
\end{aligned}
$$

the variance of estimate ***increases*** as the sample size *n* increases. **unacceptable**

3.3 Regression Methods

- Linear and Quadratic Trends in Time
- Cyclical or Seasonal Trends
- Cosine Trends

3.3.1 Linear and Quadratic Trends in Time

Consider the deterministic time trend

$$\mu_t = \beta_0 + \beta_1 t$$

where the *slope* and *intercept*, β_1 and β_0 respectively, are unknown parameters.

classical least squares (or regression) method

$$\text{minimize} \quad Q(\beta_0, \beta_1) = \sum_{t=1}^{n} [Y_t - (\beta_0 + \beta_1 t)]^2$$

3.3.1 Linear and Quadratic Trends in Time

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{n} (Y_t - \bar{Y})(t - \bar{t})}{\sum_{t=1}^{n} (t - \bar{t})^2}$$

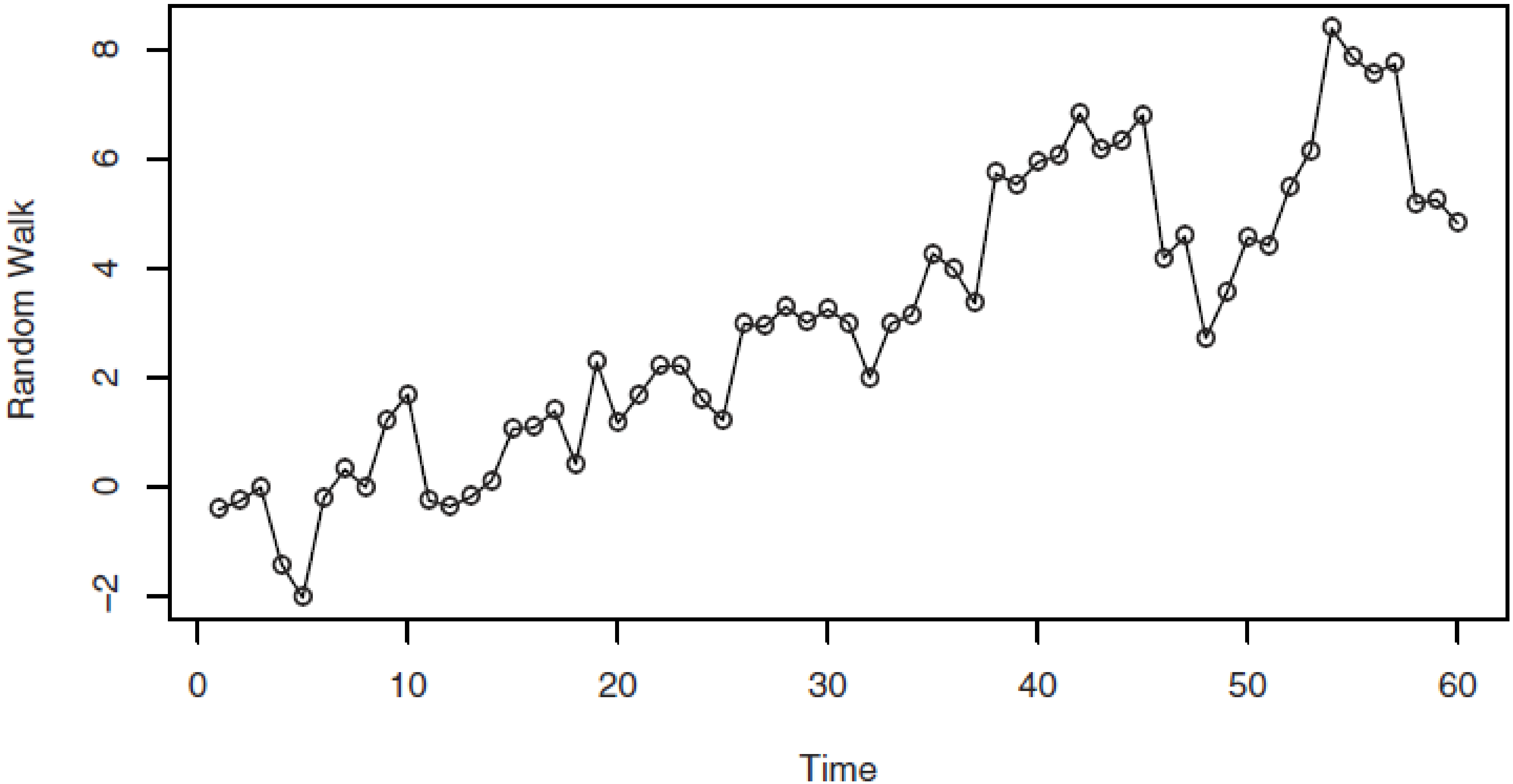
$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{t}$$

where $\bar{t} = (n + 1)/2$ is the average of $1, 2, \ldots, n$.

3.3.1 Linear and Quadratic Trends in Time

Example

Exhibit 2.1 Time Series Plot of a Random Walk



3.3.1 Linear and Quadratic Trends in Time

Example

Exhibit 3.1 Least Squares Regression Estimates for Linear Time Trend

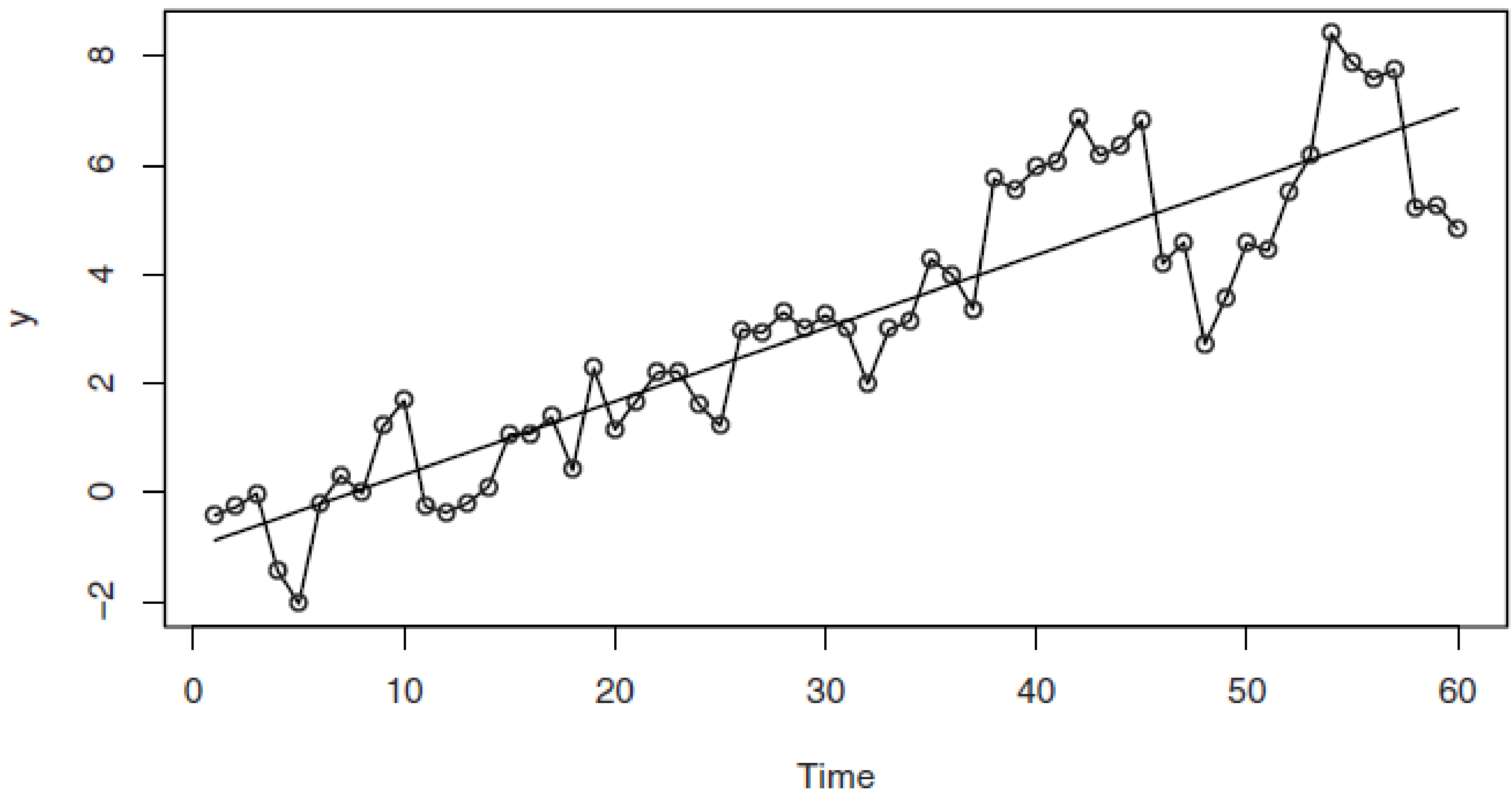
	Estimate	Std. Error	*t* value	*Pr*(>|*t*|)
Intercept	−1.008	0.2972	−3.39	0.00126
Time	0.1341	0.00848	15.82	< 0.0001

```
> data(rwalk)
> model1=lm(rwalk~time(rwalk))
> summary(model1)
```

3.3.1 Linear and Quadratic Trends in Time

Exhibit 3.2 Random Walk with Linear Time Trend

Example




```
> win.graph(width=4.875, height=2.5,pointsize=8)
> plot(rwalk,type='o',ylab='y')
> abline(model1) # add the fitted least squares line from model1
```

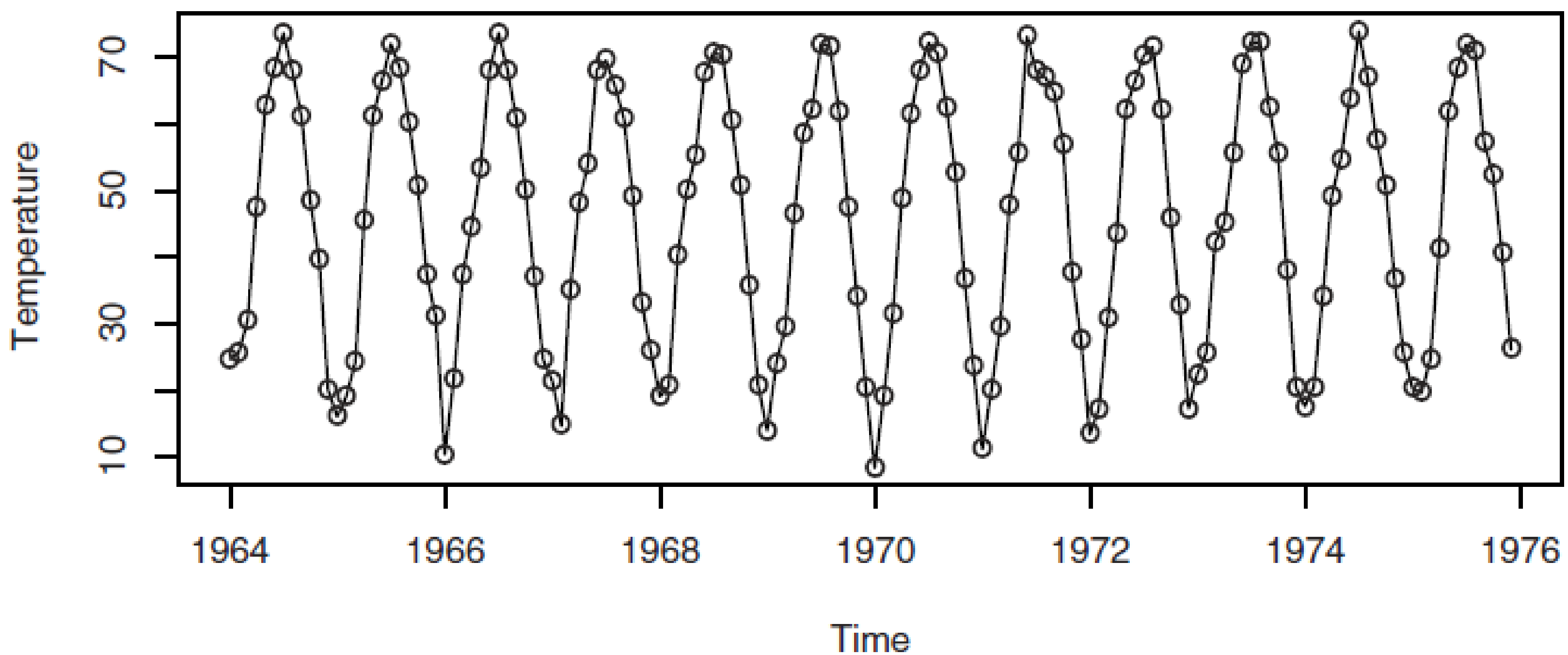
3.3.2 Cyclical or Seasonal Trends

seasonal trends

$$Y_t = \mu_t + X_t$$

where $E(X_t) = 0$ for all t.

Exhibit 1.7 Average Monthly Temperatures, Dubuque, Iowa



3.3.2 Cyclical or Seasonal Trends

general assumption for μ_t with monthly seasonal data

$$\mu_t = \begin{cases} \beta_1 & \text{for } t = 1, 13, 25, \ldots \\ \beta_2 & \text{for } t = 2, 14, 26, \ldots \\ \vdots & \\ \beta_{12} & \text{for } t = 12, 24, 36, \ldots \end{cases}$$

This is sometimes called a **seasonal means** model.

3.3.2 Cyclical or Seasonal Trends

图表 3-3 季节均值模型回归结果

	估 计	标准误差	t 值	Pr(>|t|)
1 月	16.608	0.987	16.8	<0.0001
2 月	20.650	0.987	20.9	<0.0001
3 月	32.475	0.987	32.9	<0.0001
4 月	46.525	0.987	47.1	<0.0001
5 月	58.092	0.987	58.9	<0.0001
6 月	67.500	0.987	68.4	<0.0001
7 月	71.717	0.987	72.7	<0.0001
8 月	69.333	0.987	70.2	<0.0001
9 月	61.025	0.987	61.8	<0.0001
10 月	50.975	0.987	51.6	<0.0001
11 月	36.650	0.987	37.1	<0.0001
12 月	23.642	0.987	24.0	<0.0001

```
> data(tempdub)
> month.=season(tempdub) # period added to improve table display
> model2=lm(tempdub~month.-1) # -1 removes the intercept term
> summary(model2)
```

3.3.2 Cyclical or Seasonal Trends

图表 3-4　带截距项的季节均值模型回归结果

	估　计	标准误差	t 值	Pr(>|t|)
1 月	16.608	0.987	16.83	<0.0001
2 月	4.042	1.396	2.90	0.00443
3 月	15.867	1.396	11.37	<0.0001
4 月	29.917	1.396	21.43	<0.0001
5 月	41.483	1.396	29.72	<0.0001
6 月	50.892	1.396	36.46	<0.0001
7 月	55.108	1.396	39.48	<0.0001
8 月	52.725	1.396	37.78	<0.0001
9 月	44.417	1.396	31.82	<0.0001
10 月	34.367	1.396	24.62	<0.0001
11 月	20.042	1.396	14.36	<0.0001
12 月	7.033	1.396	5.04	<0.0001

```
> model3=lm(tempdub~month.) # January is dropped automatically
> summary(model3)
```

3.3.3 Cosine Trends

Consider the cosine curve with equation

$$\mu_t = \beta\cos(2\pi f t + \Phi)$$

We call β (> 0) the *amplitude*, f the *frequency*, and Φ

Φ the *phase* of the curve.

reparameterizes

$$\beta\cos(2\pi f t + \Phi) = \beta_1\cos(2\pi f t) + \beta_2\sin(2\pi f t)$$

where $\beta = \sqrt{\beta_1^2 + \beta_2^2}$, $\Phi = \operatorname{atan}(-\beta_2/\beta_1)$

and, conversely, $\beta_1 = \beta\cos(\Phi)$, $\beta_2 = \beta\sin(\Phi)$

3.3.3 Cosine Trends

The simplest such model for the trend would be expressed as

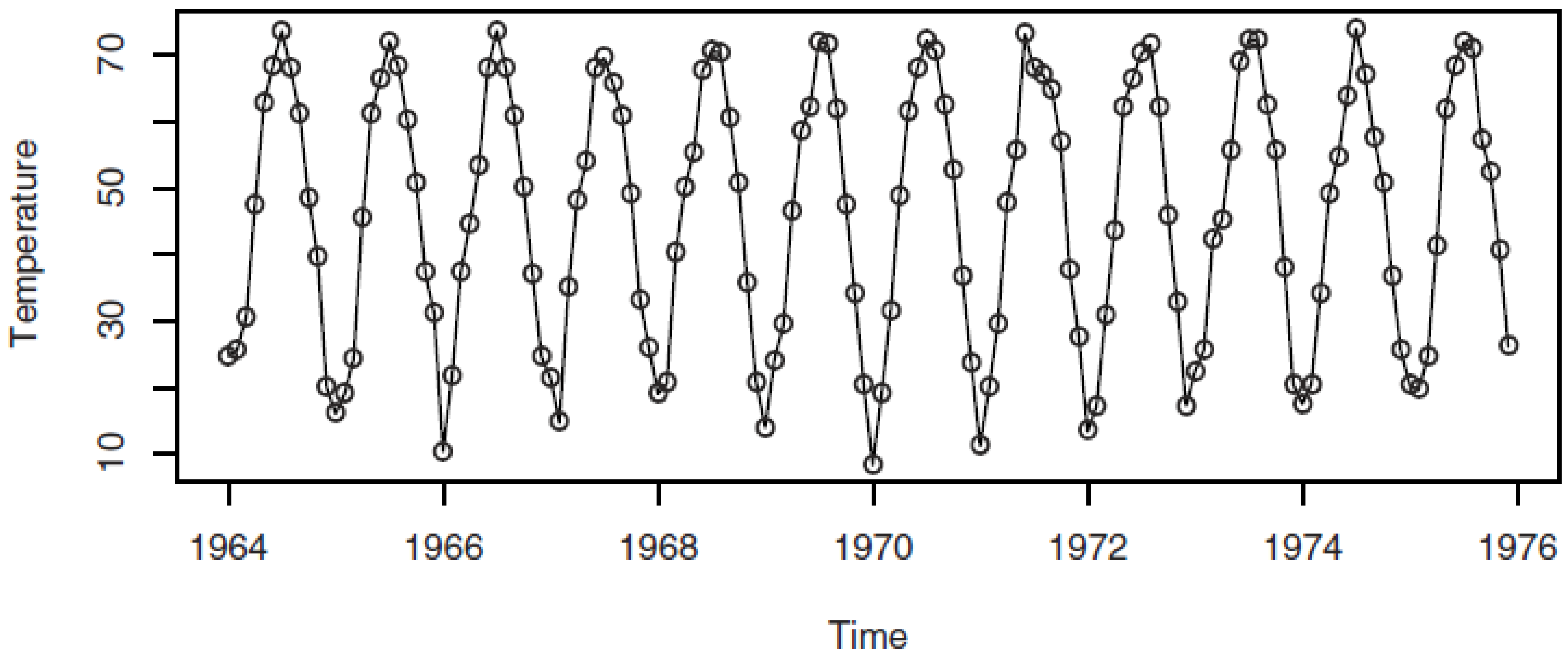
$$\mu_t = \beta_0 + \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t)$$

Here the constant term, β_0, can be meaningfully thought of as a cosine with frequency zero.

In practical, we must be careful how we **measure time**, as our choice of time measurement will **affect** the values of the **frequencies** of interest.

3.3.3 Cosine Trends

Exhibit 1.7 Average Monthly Temperatures, Dubuque, Iowa



3.3.3 Cosine Trends

Exhibit 3.5 Cosine Trend Model for Temperature Series

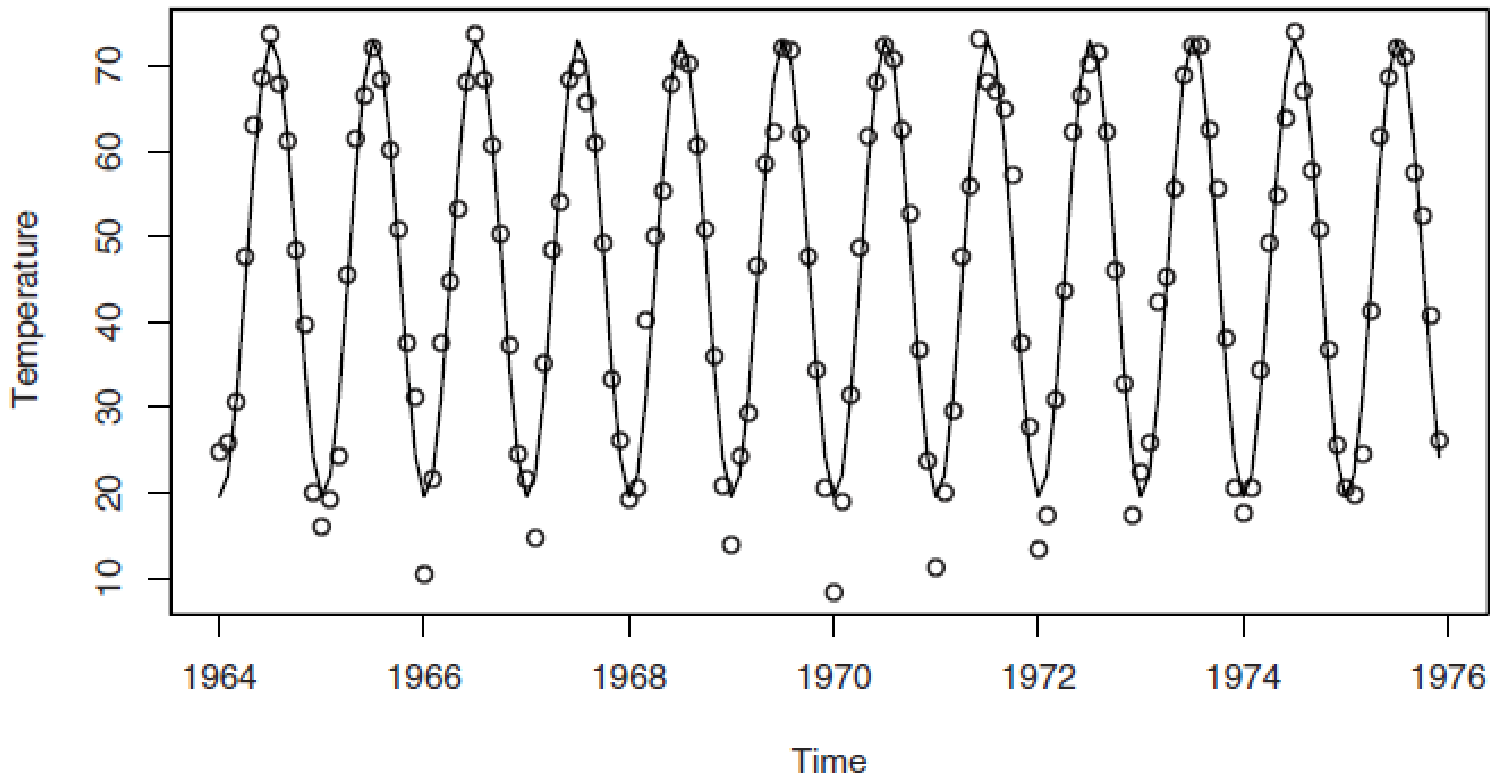
Coefficient	Estimate	Std. Error	*t*-value	*Pr*(>|*t*|)
Intercept	46.2660	0.3088	149.82	< 0.0001
$\cos(2\pi t)$	–26.7079	0.4367	–61.15	< 0.0001
$\sin(2\pi t)$	–2.1697	0.4367	–4.97	<0.0001

```
> har.=harmonic(tempdub,1)
> model4=lm(tempdub~har.)
> summary(model4)
```

In this output, time is measured in years, with 1964 as the starting value and a frequency of 1 per year.

3.3.3 Cosine Trends

Exhibit 3.6 Cosine Trend for the Temperature Series




```
> win.graph(width=4.875, height=2.5,pointsize=8)
> plot(ts(fitted(model4),freq=12,start=c(1964,1)),
   ylab='Temperature',type='l',
> ylim=range(c(fitted(model4),tempdub))); points(tempdub)
> # ylim ensures that the y axis range fits the raw data and the
   fitted values
```

3.4 Reliability and Efficiency of Regression Estimates

We assume that the series is represented as $Y_t = \mu_t + X_t$, where μ_t is a deterministic trend of the kind considered above and $\{X_t\}$ is a zero-mean stationary process with autocovariance and autocorrelation functions γ_k and ρ_k, respectively.

3.4 Reliability and Efficiency of Regression Estimates

the seasonal means

$$\hat{\beta}_j = \frac{1}{N}\sum_{i=0}^{N-1} Y_{j+12i}$$

$$Var(\hat{\beta}_j) = \frac{\gamma_0}{N}\left[1 + 2\sum_{k=1}^{N-1}\left(1 - \frac{k}{N}\right)\rho_{12k}\right]$$

This is similar to the results of **estimation of constant mean.**

3.4 Reliability and Efficiency of Regression Estimates

cosine trends

$$Y_t = \beta_0 + \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t)$$

$$\hat{\beta}_1 = \frac{2}{n} \sum_{t=1}^{n} \left[\cos\left(\frac{2\pi m t}{n}\right) Y_t \right],$$

$$f = m/n$$

$$\hat{\beta}_2 = \frac{2}{n} \sum_{t=1}^{n} \left[\sin\left(\frac{2\pi m t}{n}\right) Y_t \right]$$

the correlations between the time series $\{Y_t\}$ and the cosine and sine waves with frequency m/n.

3.4 Reliability and Efficiency of Regression Estimates

the fact that $\sum_{t=1}^{n}[\cos(2\pi mt/n)]^2 = n/2$

$$Var(\hat{\beta}_1) = \frac{2\gamma_0}{n}\left[1 + \frac{4}{n}\sum_{s=2}^{n}\sum_{t=1}^{s-1}\cos\left(\frac{2\pi mt}{n}\right)\cos\left(\frac{2\pi ms}{n}\right)\rho_{s-t}\right]$$

If $\{X_t\}$ is white noise, we get just $2\gamma_0/n$.

If $\rho_1 \neq 0$, $\rho_k = 0$ for $k > 1$, and $m/n = 1/12$,

$$Var(\hat{\beta}_1) = \frac{2\gamma_0}{n}\left[1 + \frac{4\rho_1}{n}\sum_{t=1}^{n-1}\cos\left(\frac{\pi t}{6}\right)\cos\left(\frac{\pi t + 1}{6}\right)\right]$$

3.4 Reliability and Efficiency of Regression Estimates

n	$Var(\hat{\beta}_1)$
25	$\left(\frac{2\gamma_0}{n}\right)(1 + 1.71\rho_1)$
50	$\left(\frac{2\gamma_0}{n}\right)(1 + 1.75\rho_1)$
500	$\left(\frac{2\gamma_0}{n}\right)(1 + 1.73\rho_1)$
∞	$\left(\frac{2\gamma_0}{n}\right)\left(1 + 2\rho_1\cos\left(\frac{\pi}{6}\right)\right) = \left(\frac{2\gamma_0}{n}\right)\boxed{(1 + 1.732\rho_1)}$

If $\rho_1 = -0.4, \quad 1+1.732(-0.4) = 0.307$

How to **compare** the models? seasonal means or cosine trends?

3.4 Reliability and Efficiency of Regression Estimates

The parameters themselves are not directly comparable, but we can compare the estimates of the trend at **comparable time points**.

Why?

Consider the two estimates for the trend in January

seasonal means

$$Var(\hat{\beta}_j) = \frac{\gamma_0}{N}\left[1 + 2\sum_{k=1}^{N-1}\left(1 - \frac{k}{N}\right)\rho_{12k}\right] \qquad \text{for } j = 1$$

3.4 Reliability and Efficiency of Regression Estimates

cosine trend model

$$\hat{\mu}_1 = \hat{\beta}_0 + \hat{\beta}_1 \cos\left(\frac{2\pi}{12}\right) + \hat{\beta}_2 \sin\left(\frac{2\pi}{12}\right)$$

$$Var(\hat{\mu}_1) = Var(\hat{\beta}_0) + Var(\hat{\beta}_1)\left[\cos\left(\frac{2\pi}{12}\right)\right]^2 + Var(\hat{\beta}_2)\left[\sin\left(\frac{2\pi}{12}\right)\right]^2$$

Case 1: white noise.

$$Var(\hat{\mu}_1) = \frac{\gamma_0}{n}\left\{1 + 2\left[\cos\left(\frac{\pi}{6}\right)\right]^2 + 2\left[\sin\left(\frac{\pi}{6}\right)\right]^2\right\} = 3\frac{\gamma_0}{n}$$

3.4 Reliability and Efficiency of Regression Estimates

the ratio of the standard deviation in the **cosine model** to that in the **seasonal means** model:

$$\sqrt{\frac{3\gamma_0/n}{\gamma_0/N}} = \sqrt{\frac{3N}{n}} \qquad n = 12N$$

Thus, in the cosine model, we estimate the January effect with a standard deviation that is **only half as large as** it would be if we estimated with a seasonal means model—a substantial gain.

simpler model with smaller variance

3.4 Reliability and Efficiency of Regression Estimates

Case 2: $\rho_1 \neq 0$ but $\rho_k = 0$ for $k > 1$

seasonal means

$$Var(\hat{\beta}_j) = \frac{\gamma_0}{N}$$

cosine trend model

$$Var(\hat{\mu}_1) = \frac{\gamma_0}{n}\left\{1 + 2\rho_1 + 2\left[1 + 2\rho_1 \cos\left(\frac{2\pi}{12}\right)\right]\right\}$$

$$= \frac{\gamma_0}{n}\left\{3 + 2\rho_1\left[1 + 2\cos\left(\frac{\pi}{6}\right)\right]\right\}$$

3.4 Reliability and Efficiency of Regression Estimates

If $\rho_1 = -0.4$, then we have $0.814\gamma_0/n$,

the ratio of the standard deviation in the **cosine model** to that in the **seasonal means** model:

$$\sqrt{\left[\frac{(0.814\gamma_0)/n}{\gamma_0/N}\right]} = \sqrt{\frac{0.814N}{n}} = 0.26$$

a very substantial reduction indeed!

The cosine trend plus white noise model is the correct model?

3.4 Reliability and Efficiency of Regression Estimates

linear time trends

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{n} (t - \bar{t})Y_t}{\sum_{t=1}^{n} (t - \bar{t})^2}$$

$$\boxed{\sum_{t=1}^{n} (t - \bar{t})^2 = n(n^2 - 1)/12}$$

$$Var(\hat{\beta}_1) = \frac{12\gamma_0}{n(n^2 - 1)}\left[1 + \frac{24}{n(n^2 - 1)} \sum_{s=2}^{n} \sum_{t=1}^{s-1} (t - \bar{t})(s - \bar{t})\rho_{s-t}\right]$$

3.4 Reliability and Efficiency of Regression Estimates

Case: $\rho_1 \neq 0$ but $\rho_k = 0$ for $k > 1$

$$Var(\hat{\beta}_1) = \frac{12\gamma_0}{n(n^2-1)}\left[1 + 2\rho_1\left(1 - \frac{3}{n}\right)\right]$$

$$= \frac{12\gamma_0(1+2\rho_1)}{n(n^2-1)}$$

If $\rho_1 = -0.4$, then $1 + 2\rho_1 = 0.2$, and then the variance of $\hat{\beta}_1$ is only 20% of what it would be if $\{X_t\}$ were white noise.

3.5 Interpreting Regression Output

residual standard deviation

$$s = \sqrt{\frac{1}{n-p} \sum_{t=1}^{n} (Y_t - \hat{\mu}_t)^2}$$

$\hat{\mu}_t$: the estimated trend for μ_t

p the number of parameters estimated in μ_t

$n-p$: *degrees of freedom* for *s*

difficult to interpret

The value of *s* gives an absolute measure of the goodness of fit of the estimated trend— **the smaller the value of *s*, the better the fit.**

3.5 Interpreting Regression Output

coefficient of determination (*R***-squared)**

- the **square of the sample correlation coefficient** between the observed series and the estimated trend.

- the **fraction of the variation** in the series that is explained by the estimated trend.

unitless measure

3.5 Interpreting Regression Output

Exhibit 3.7 Regression Output for Linear Trend Fit of Random Walk

	Estimate	Std. Error	*t*-value	*Pr*(>|*t*|)
Intercept	−1.007888	0.297245	−3.39	0.00126
Time	0.134087	0.008475	15.82	< 0.0001

Residual standard error	1.137	with 58 degrees of freedom
Multiple *R*-Squared	0.812	
Adjusted *R*-squared	0.809	
F-statistic	250.3	with 1 and 58 df; *p*-value < 0.0001

```
> model1=lm(rwalk~time(rwalk))
> summary(model1)
```

The adjusted R-squared is useful for comparing models with **different** numbers of parameters.

3.5 Interpreting Regression Output

standard deviations of coefficients

The standard deviations of the coefficients labeled **Std. Error** on the output need to be interpreted carefully. They are appropriate only when **the stochastic component is white noise**—the usual regression assumption.

The important point is that these standard deviations assume a white noise stochastic component **that will rarely be true for time series**.

3.5 Interpreting Regression Output

standard deviations of coefficients

$$Var(\hat{\beta}_1) = \boxed{\frac{12\gamma_0}{n(n^2-1)}}\left[1 + \frac{24}{n(n^2-1)}\sum_{s=2}^{n}\sum_{t=1}^{s-1}(t-\bar{t})(s-\bar{t})\rho_{s-t}\right]$$

γ_0 estimated by s^2

$$0.008475 = \sqrt{\frac{12(1.137)^2}{60(60^2-1)}}$$

3.5 Interpreting Regression Output

Exhibit 3.7 Regression Output for Linear Trend Fit of Random Walk

	Estimate	Std. Error	*t*-value	*Pr*(>|*t*|)
Intercept	−1.007888	0.297245	−3.39	0.00126
Time	0.134087	$\boxed{0.008475}$	15.82	< 0.0001

Residual standard error	1.137	with 58 degrees of freedom
Multiple *R*-Squared	0.812	
Adjusted *R*-squared	0.809	
F-statistic	250.3	with 1 and 58 df; *p*-value < 0.0001

```
> model1=lm(rwalk~time(rwalk))
> summary(model1)
```

3.5 Interpreting Regression Output

t-values or *t*-ratios

The *t*-values or are just the estimated regression coefficients, each divided by their respective standard errors. **If the stochastic component is normally distributed white noise**, then these ratios provide appropriate **test statistics for checking the significance** of the regression coefficients.

3.5 Interpreting Regression Output

Exhibit 3.7 Regression Output for Linear Trend Fit of Random Walk

	Estimate	Std. Error	*t*-value	*Pr*(>|*t*|)
Intercept	−1.007888	0.297245	−3.39	0.00126
Time	0.134087	0.008475	$\boxed{15.82}$	$\boxed{< 0.0001}$

Residual standard error	1.137	with 58 degrees of freedom
Multiple *R*-Squared	0.812	
Adjusted *R*-squared	0.809	
***F*-statistic**	250.3	with 1 and 58 df; *p*-value < 0.0001

```
> model1=lm(rwalk~time(rwalk))
> summary(model1)
```

3.6 Residual Analysis

residual

$$\hat{X}_t = Y_t - \hat{\mu}_t$$

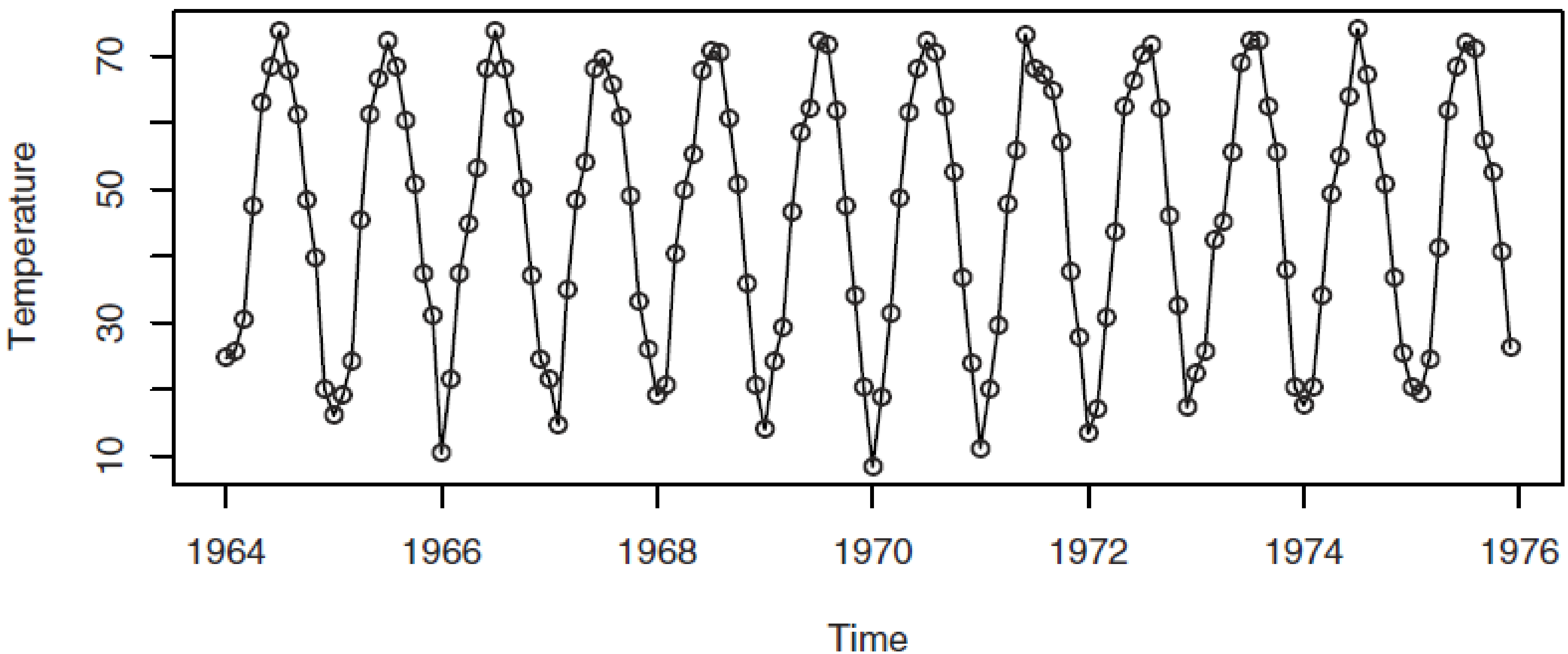
standardized residuals $\hat{X}_t / s$

standard deviation

If the trend model is reasonably **correct**, then the residuals should **behave roughly like the true stochastic component**, and various assumptions about the stochastic component can be assessed by **looking at the residuals**.

3.6 Residual Analysis

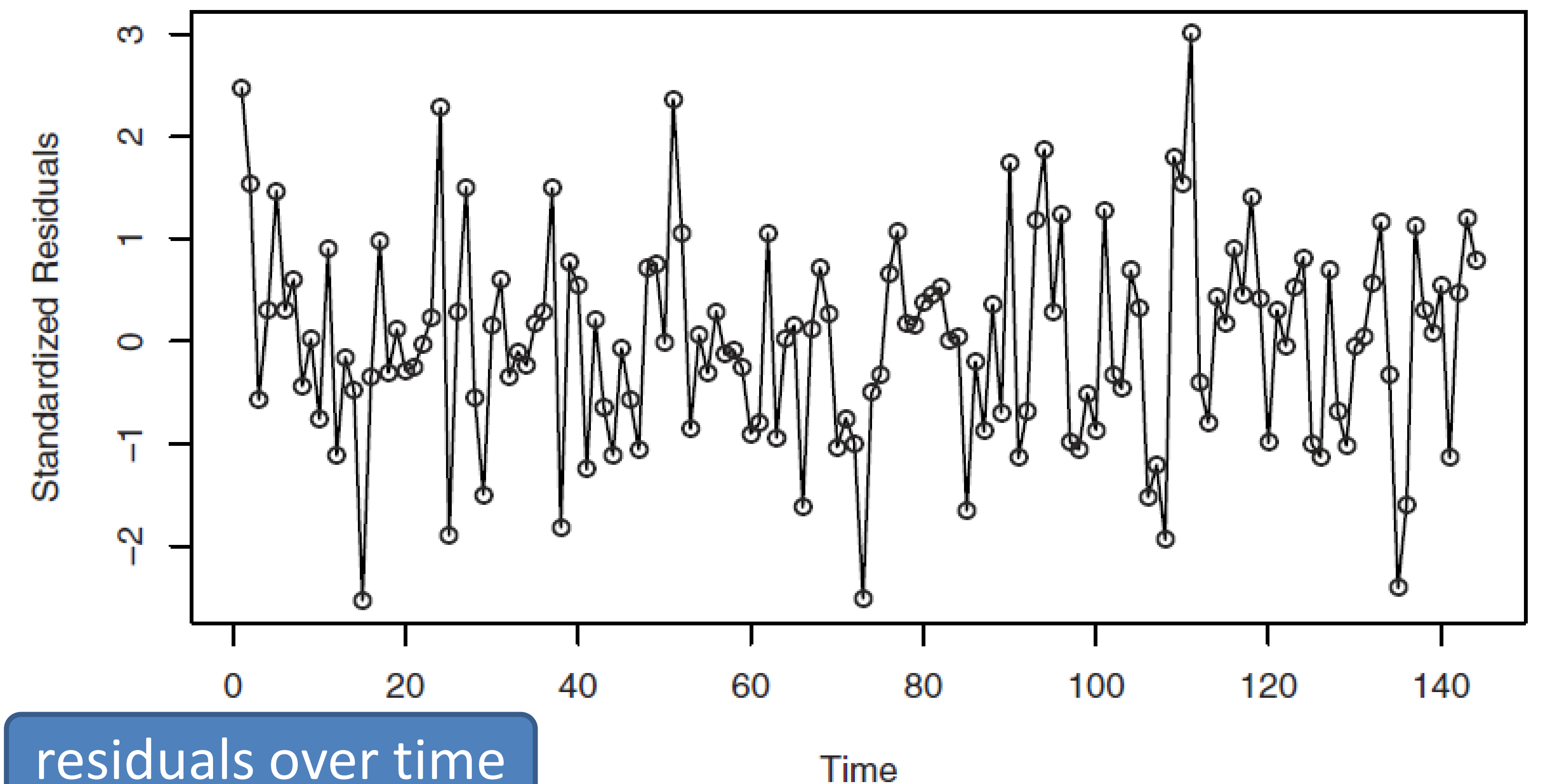
Exhibit 1.7 Average Monthly Temperatures, Dubuque, Iowa




```
> win.graph(width=4.875, height=2.5,pointsize=8)
> data(tempdub); plot(tempdub,ylab='Temperature',type='o')
```

3.6 Residual Analysis

Exhibit 3.8 **Residuals versus Time for Temperature Seasonal Means**



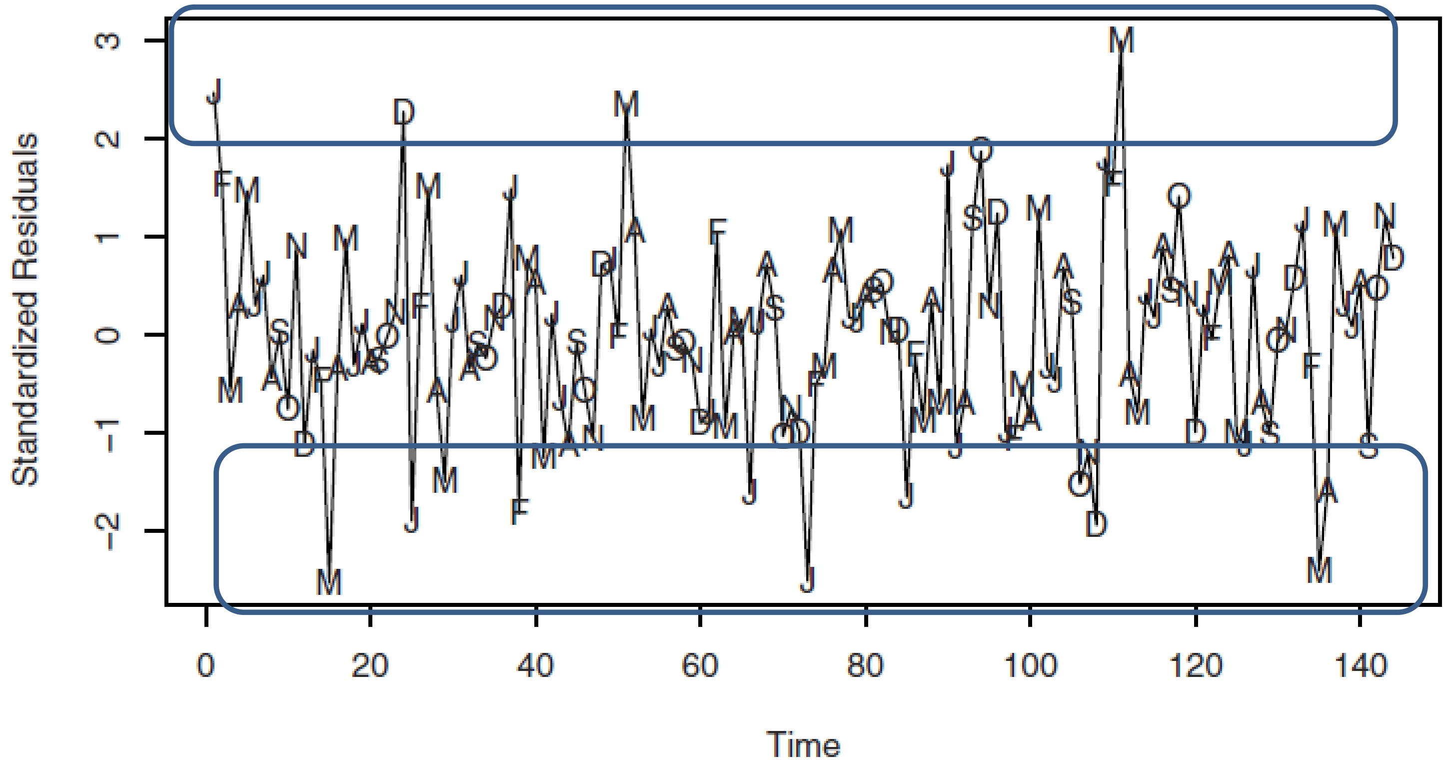
residuals over time

```
> plot(y=rstudent(model3),x=as.vector(time(tempdub)),
   xlab='Time',ylab='Standardized Residuals',type='o')
```

no striking departures from randomness apparent

3.6 Residual Analysis

Exhibit 3.9 Residuals versus Time with Seasonal Plotting Symbols

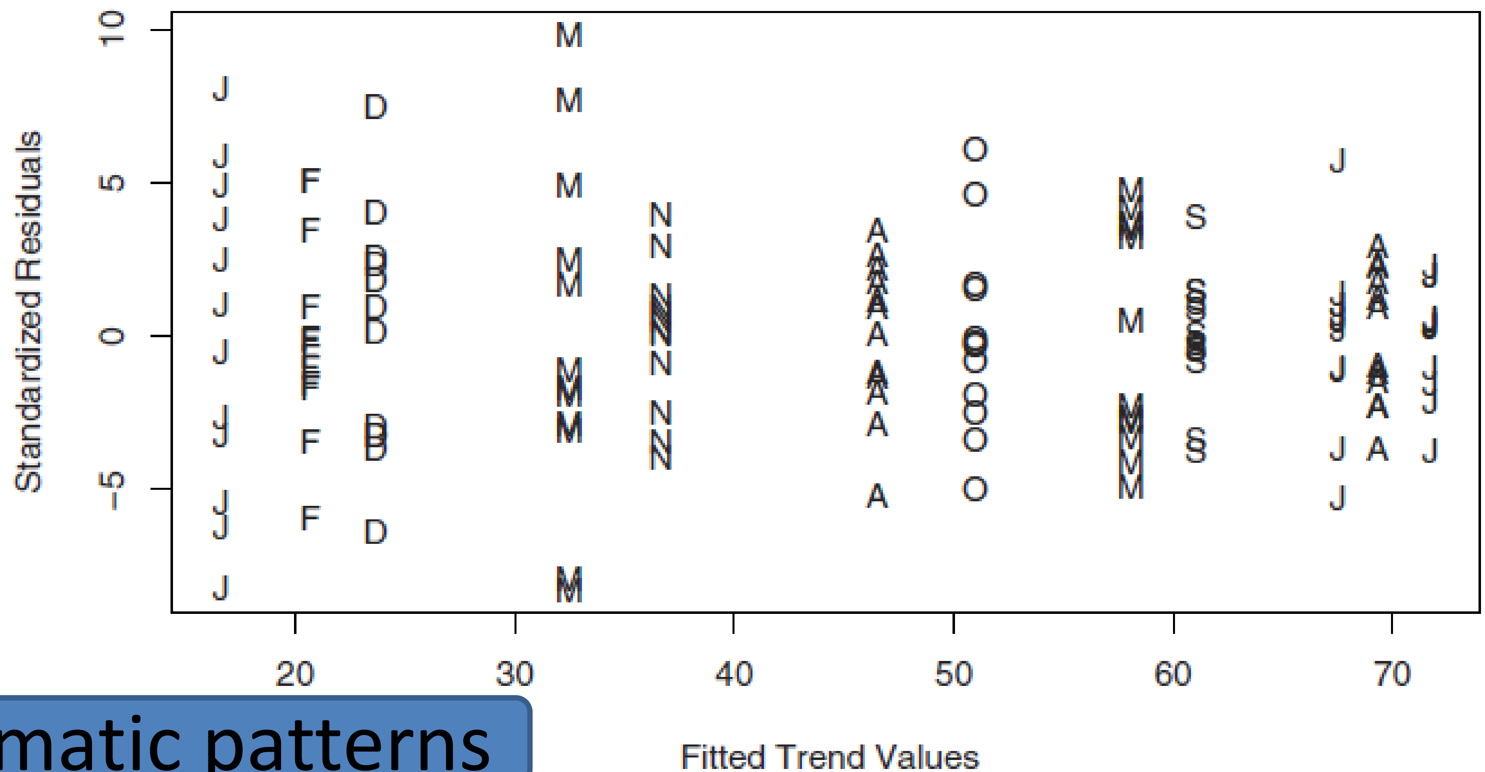



```
> plot(y=rstudent(model3),x=as.vector(time(tempdub)),xlab='Time',
> ylab='Standardized Residuals',type='l')
> points(y=rstudent(model3),x=as.vector(time(tempdub)),
   pch=as.vector(season(tempdub)))
```

no apparent patterns relating to different months

3.6 Residual Analysis

Exhibit 3.10 Standardized Residuals versus Fitted Values for the Temperature Seasonal Means Model

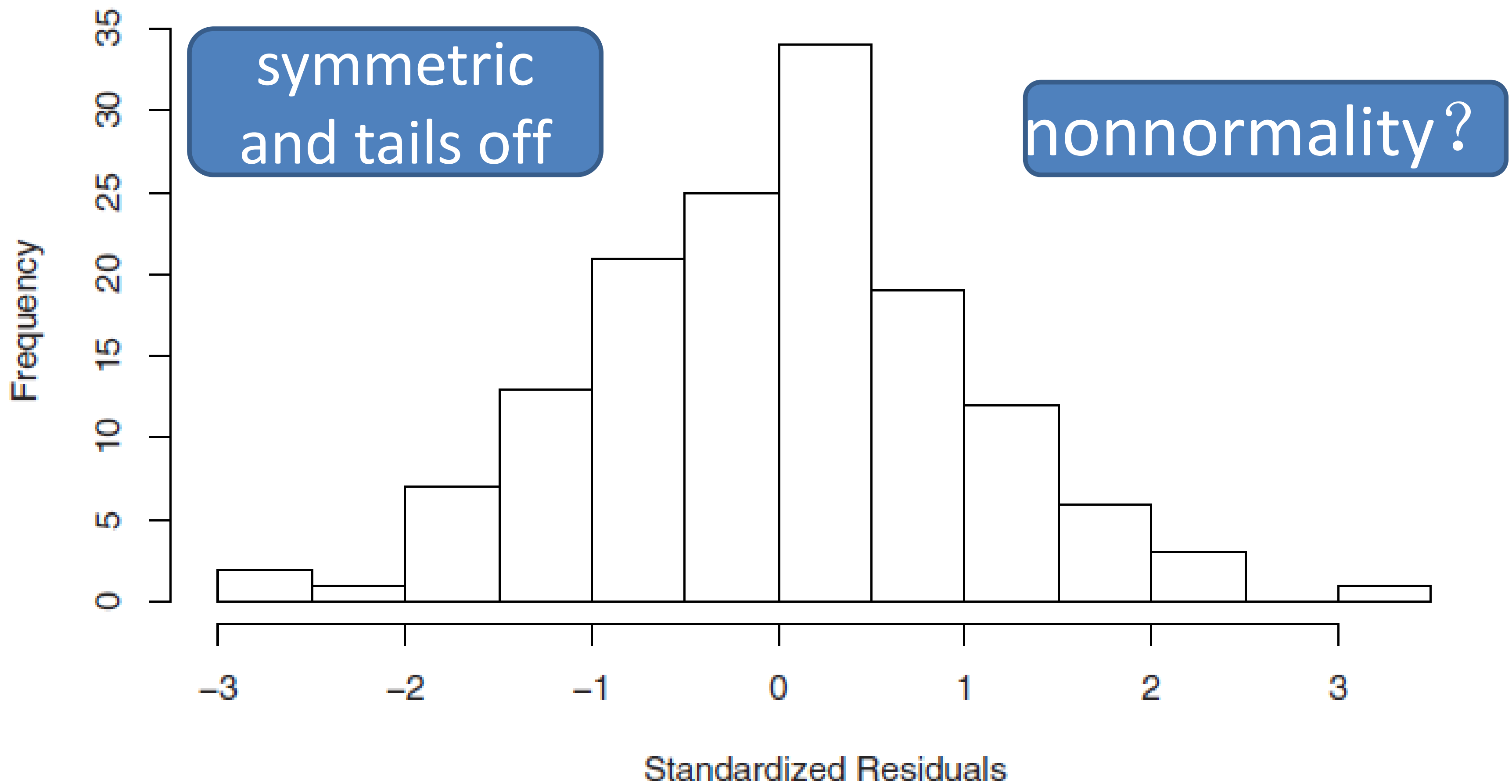


no dramatic patterns

```
> plot(y=rstudent(model3),x=as.vector(fitted(model3)),
   xlab='Fitted Trend Values',
> ylab='Standardized Residuals',type='n')
> points(y=rstudent(model3),x=as.vector(fitted(model3)),
   pch=as.vector(season(tempdub)))
```

3.6 Residual Analysis

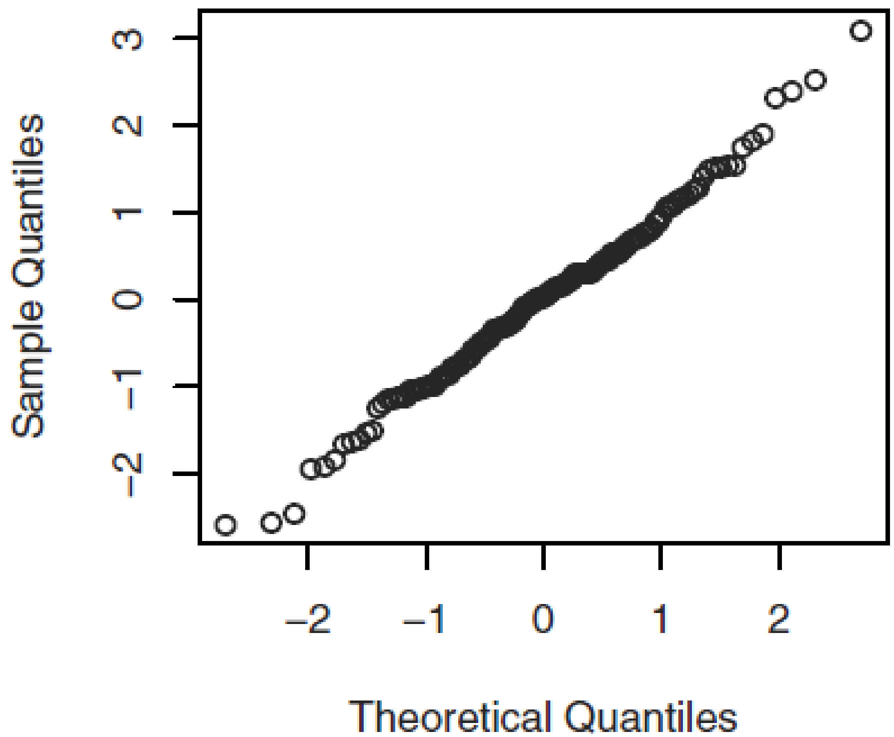
Exhibit 3.11 Histogram of Standardized Residuals from Seasonal Means Model




```
> hist(rstudent(model3),xlab='Standardized Residuals')
```

3.6 Residual Analysis

Exhibit 3.12 Q-Q Plot: Standardized Residuals of Seasonal Means Model



nonnormality？

```
> win.graph(width=2.5,height=2.5,pointsize=8)
> qqnorm(rstudent(model3))
```

Shapiro-Wilk test: *p*-value of 0.6954

3.6 Residual Analysis

Independence?

runs test

Runs above or below their **median** are counted. A **small** number of runs would indicate that neighboring residuals are **positively** dependent. On the other hand, too **many** runs would indicate neighboring residuals are **negatively** dependent. So **either too few or too many** runs lead us to **reject independence**.

observed runs = 65, expected runs = 72.875, *p*-value = 0.216, cannot reject independence

3.6 Residual Analysis

Sample Autocorrelation Function

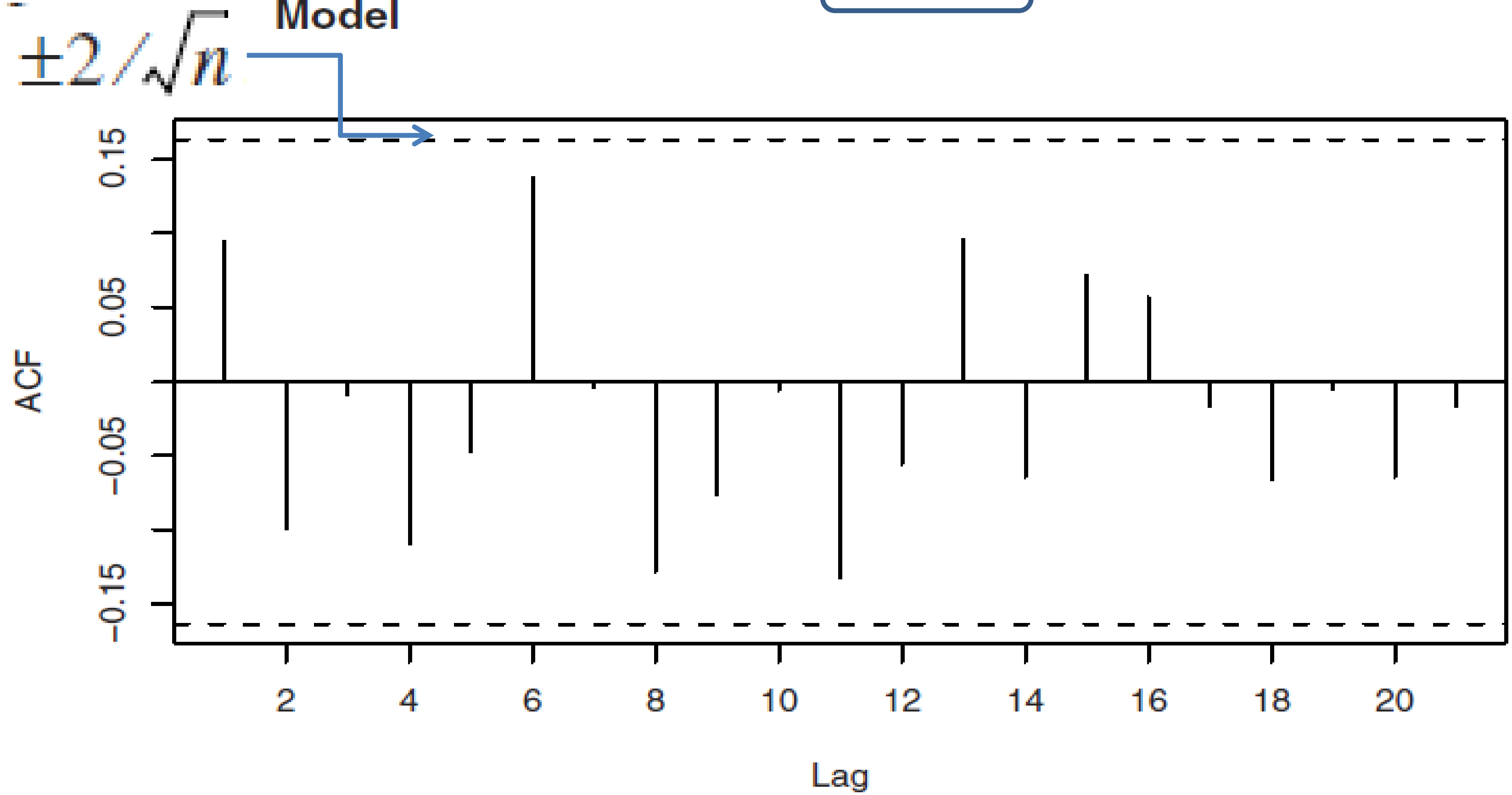
examining dependence

$$r_k = \frac{\sum_{t=k+1}^{n} (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2} \qquad \text{for } k = 1, 2, \ldots$$

A plot of r_k versus lag k is often called a **correlogram**.

3.6 Residual Analysis

Exhibit 3.13 Sample Autocorrelation of Residuals of Seasonal Means Model

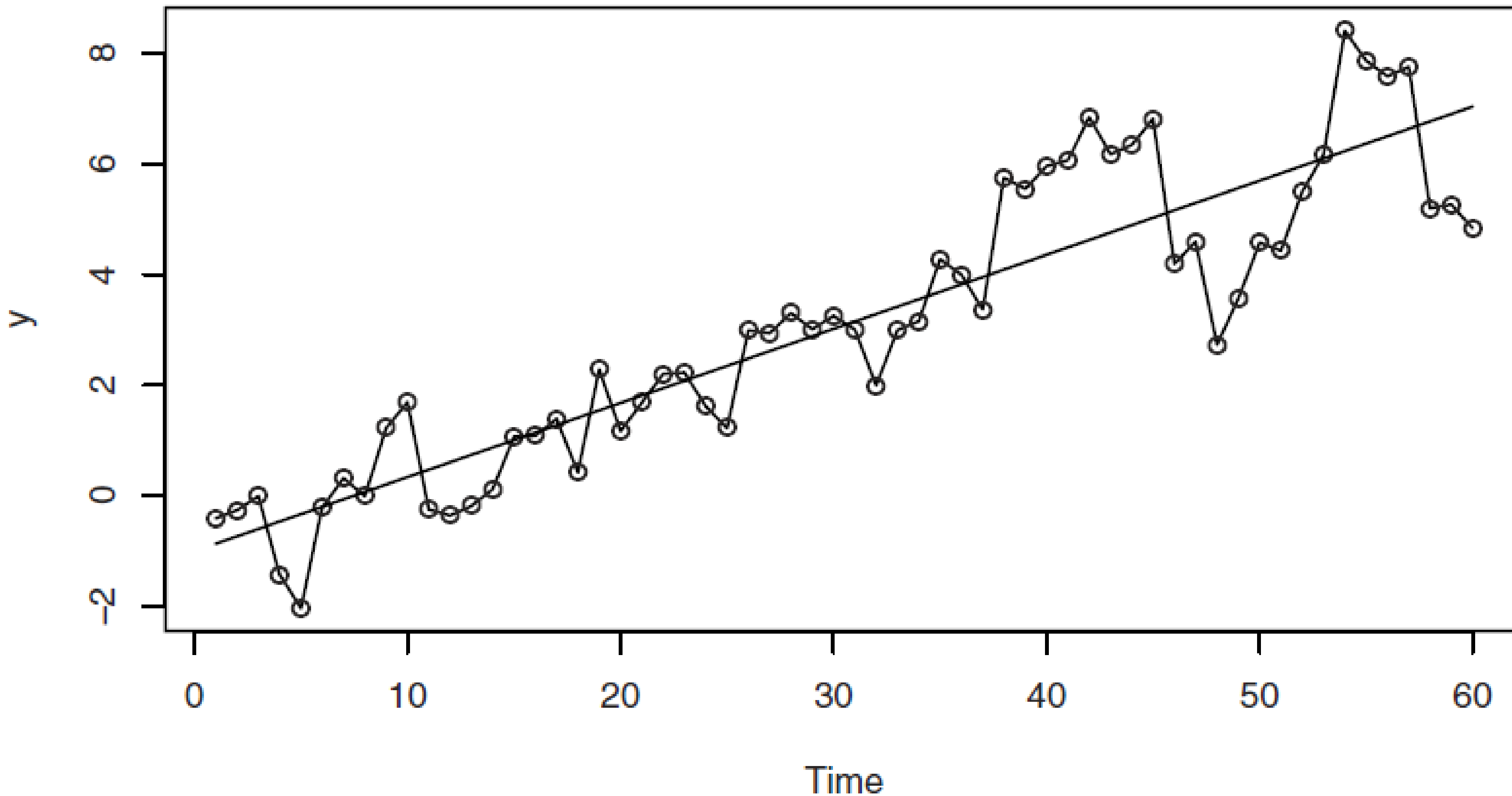



```
> win.graph(width=4.875,height=3,pointsize=8)
> acf(rstudent(model3))
```

none of the hypotheses $\rho_k = 0$ can be rejected

3.6 Residual Analysis

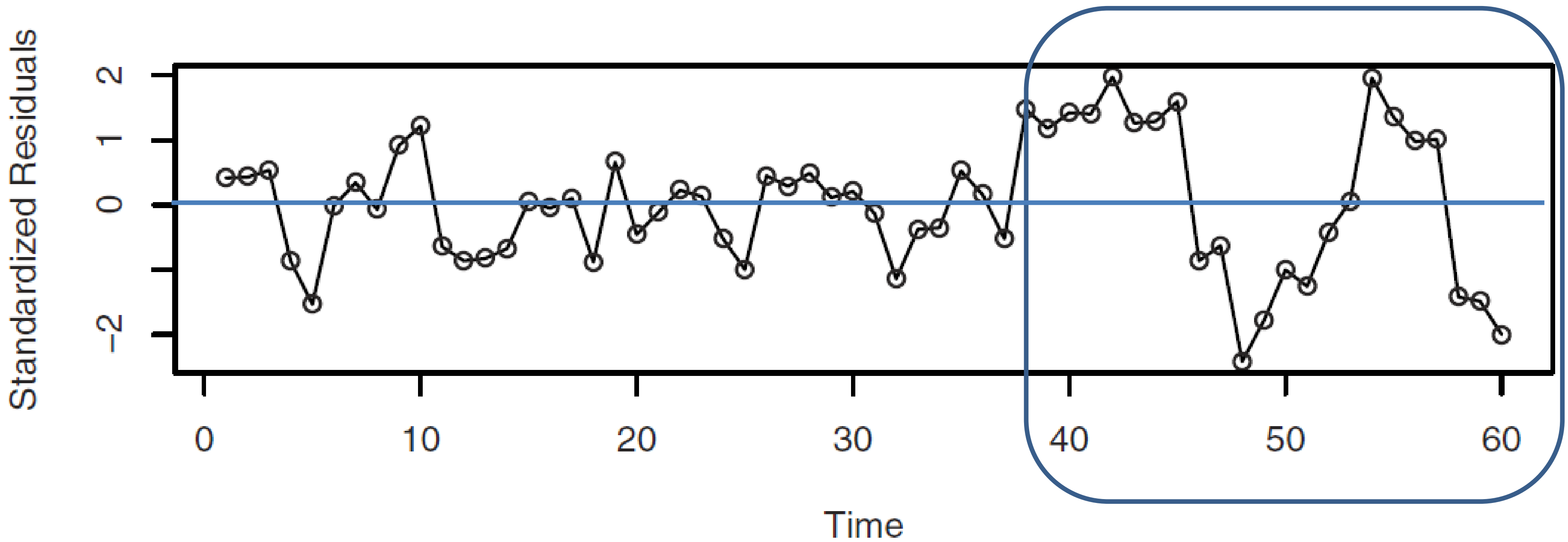
Exhibit 3.2 Random Walk with Linear Time Trend




```
> win.graph(width=4.875, height=2.5,pointsize=8)
> plot(rwalk,type='o',ylab='y')
> abline(model1) # add the fitted least squares line from model1
```

3.6 Residual Analysis

Exhibit 3.14 Residuals from Straight Line Fit of the Random Walk



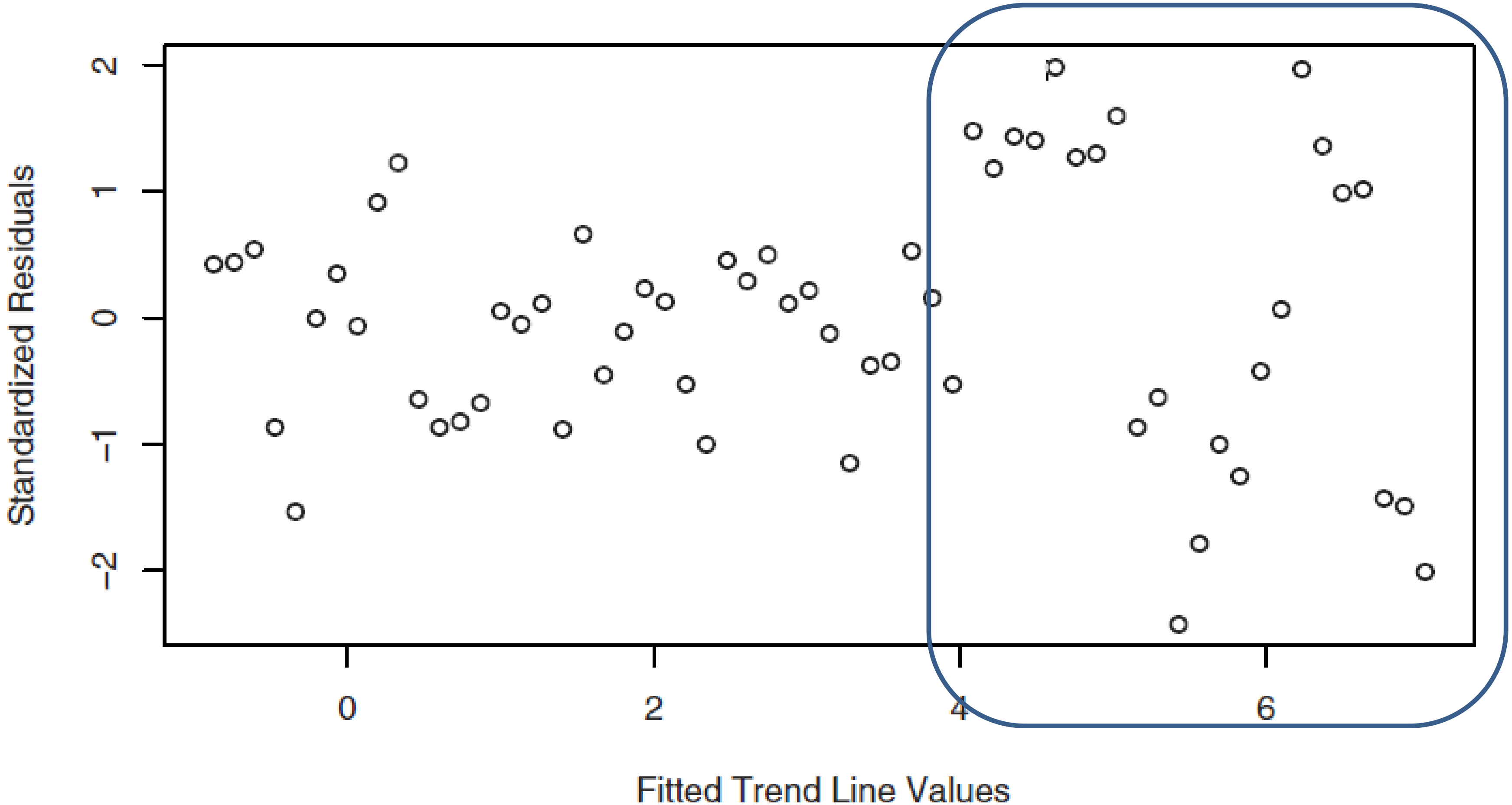

```
> plot(y=rstudent(model1),x=as.vector(time(rwalk)),
   ylab='Standardized Residuals',xlab='Time',type='o')
```

1. the residuals “hang together” too much for white noise

2.more variation in the last third of the series

3.6 Residual Analysis

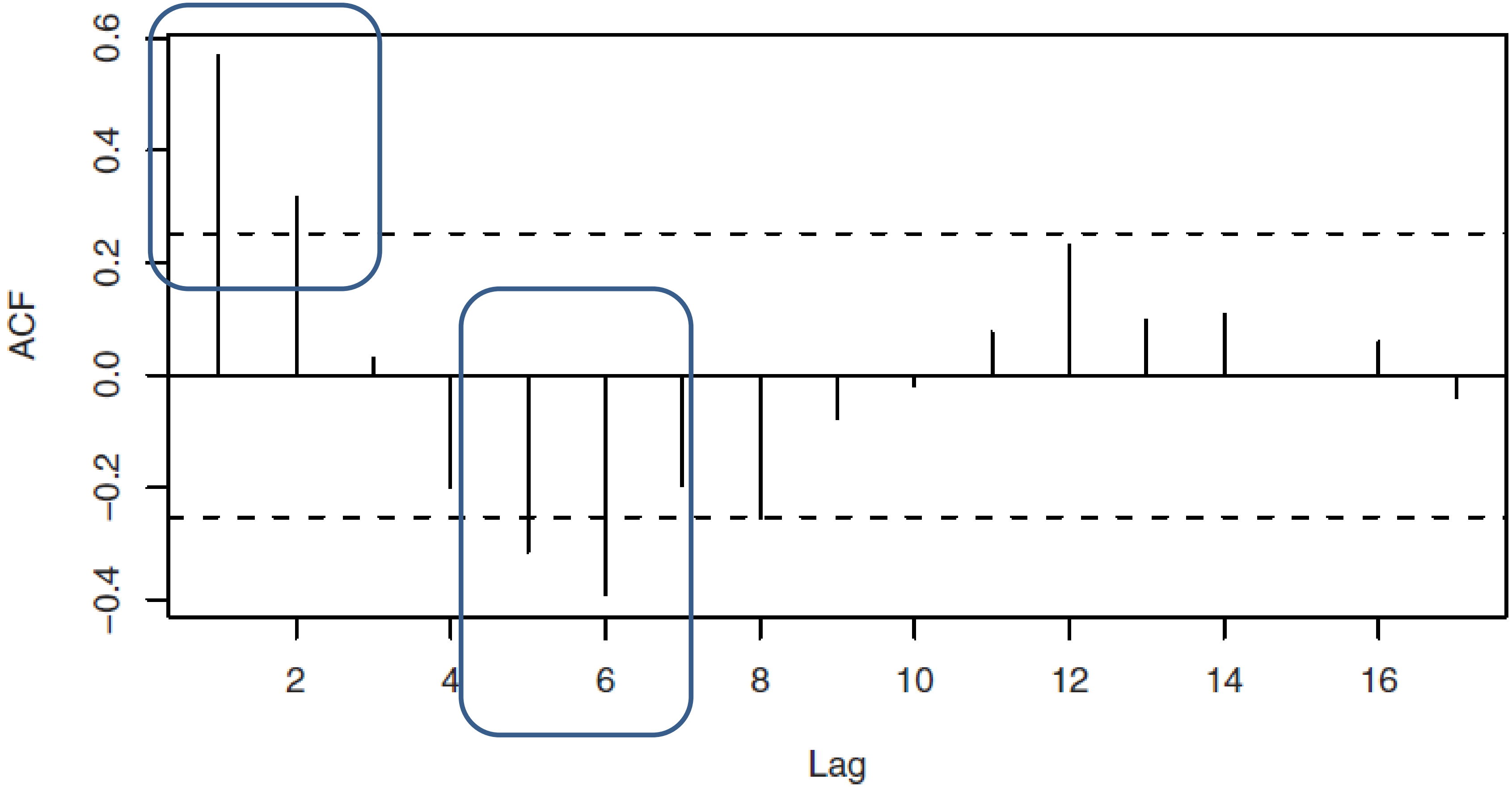
Exhibit 3.15 Residuals versus Fitted Values from Straight Line Fit



larger residuals associated with larger fitted values.

3.6 Residual Analysis

Exhibit 3.16 Sample Autocorrelation of Residuals from Straight Line Model




```
> acf(rstudent(model1))
```

This is not what we expect from a white noise process.

3.7 Summary

This chapter is concerned with describing, modeling, and estimating **deterministic trends**. Methods of estimating a constant mean were given but, more importantly, **assessment of the accuracy** of the estimates under various conditions was considered. **Regression methods** were then pursued to estimate trends that are linear or quadratic in time. Methods for modeling cyclical or seasonal trends came next, and the **reliability and efficiency** of all of these regression methods were investigated. The final section began our study of residual analysis to **investigate the quality of the fitted model**. This section also introduced the **important sample autocorrelation function**.

作业

用程序语言实现常均值模型、线性模型、二次模型、季节趋势模型、余弦趋势模型的参数估计，并使用教才中相应数据检验算法的正确性

基于上面的程序，完成3.5、3.7、3.11、3.14题