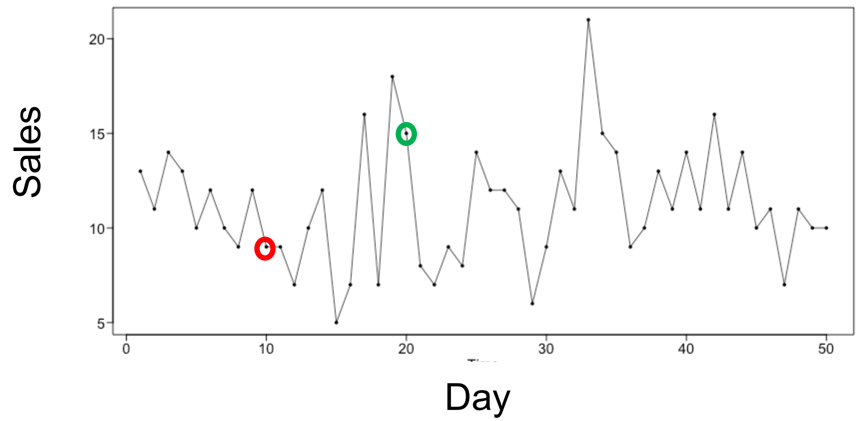
Notation and Examples

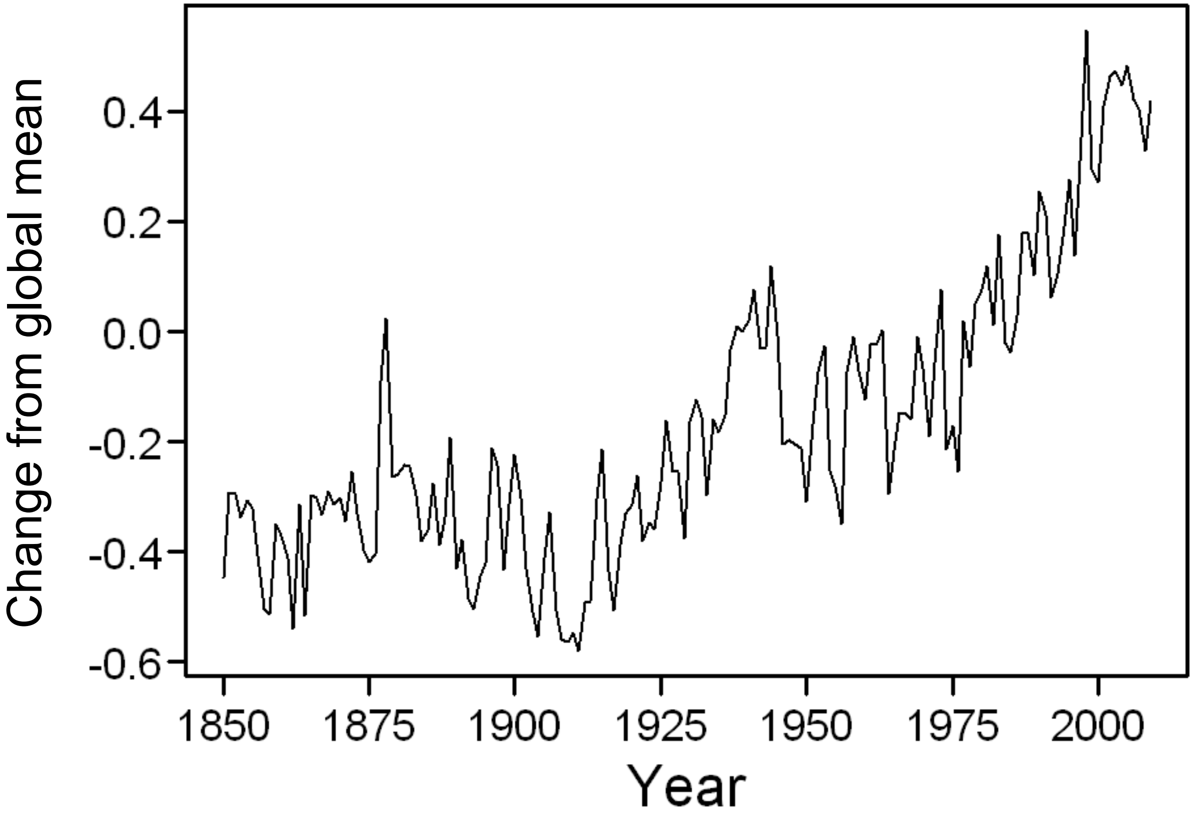
Notation



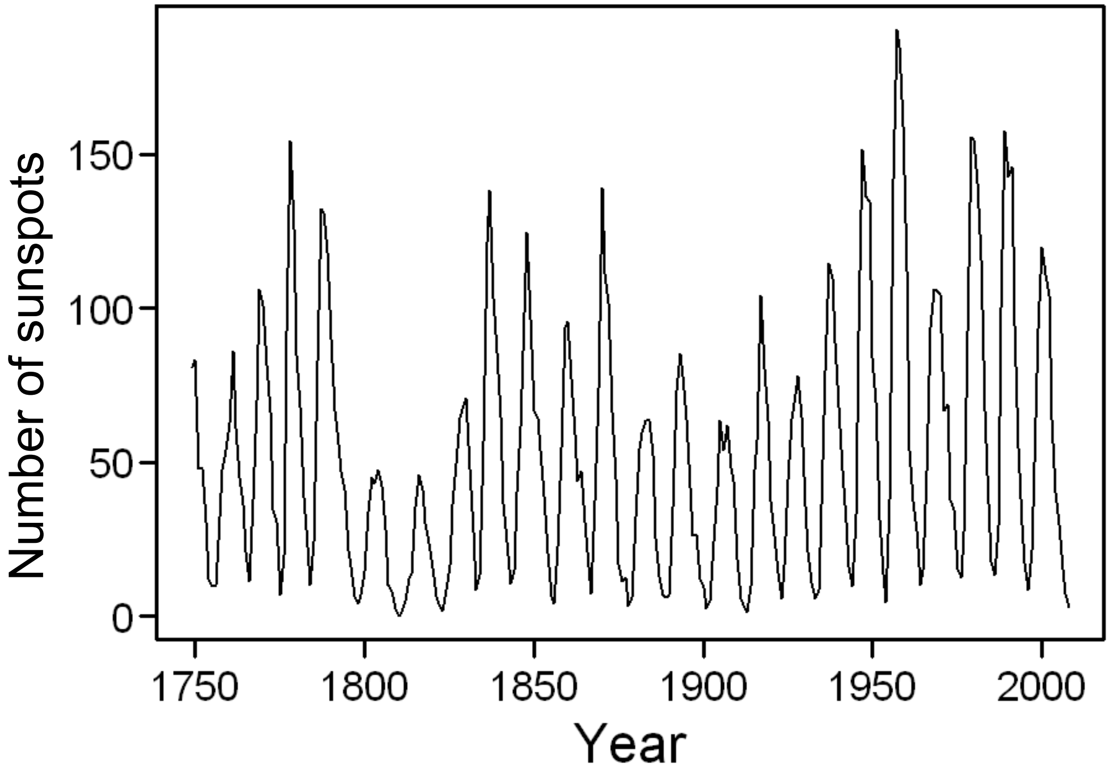
- Above is a plot of the sales of a new dishwashing detergent at a particular local Walmart. The *x*-axis is time (day) and the *y*-axis is the response: sales.
- As a matter of convention, X_t is the response, and time (day) is the explanatory variable.
- We will define X_t to be the random variable that represents sales at day *t* and will let x_t represent a particular value of X_t from the sample (known as a realization of the time series). For example… from the realization above, x_{10} = 9 and similarly x_{20} = 15.
- The collection of all $X_t s$ (one for each *t*) is known as a *stochastic process* and may also be referred to as a *time series*.

Examples of Time Series

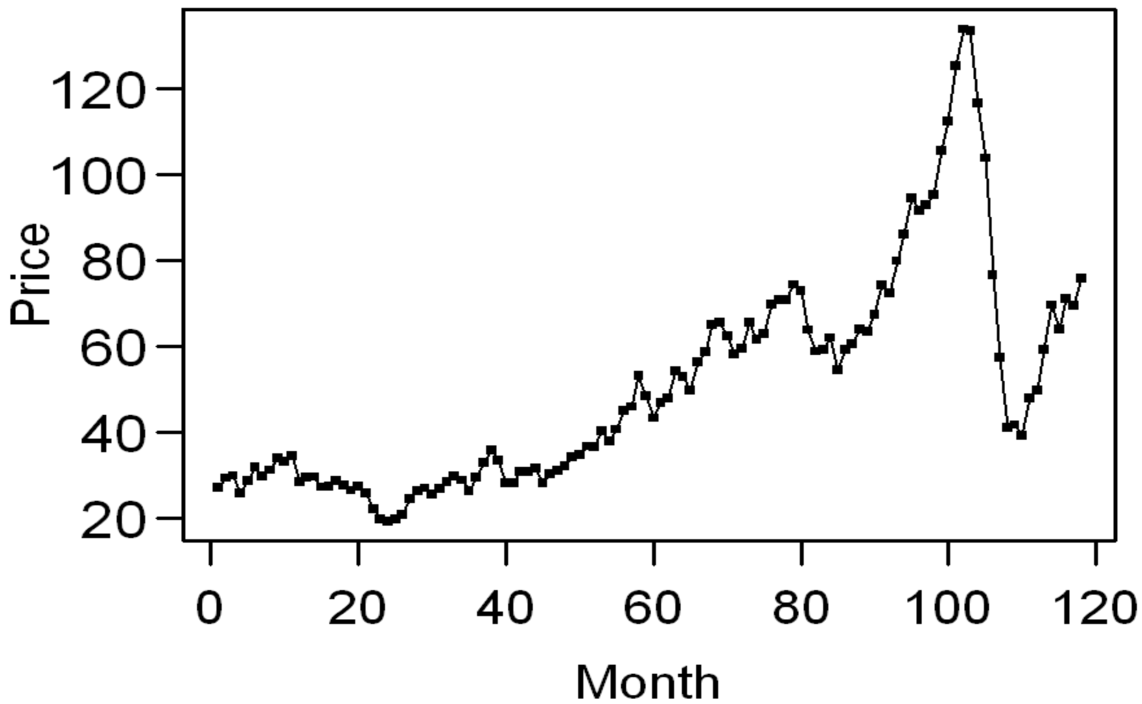
Global Temperatures



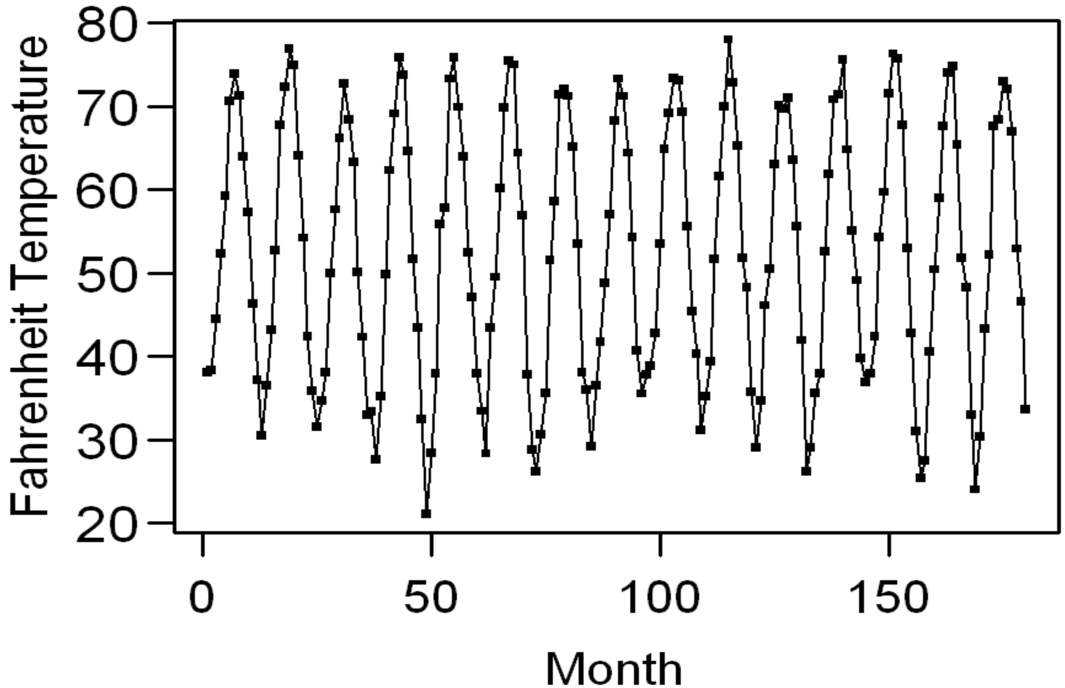
Annual Sunspot Numbers



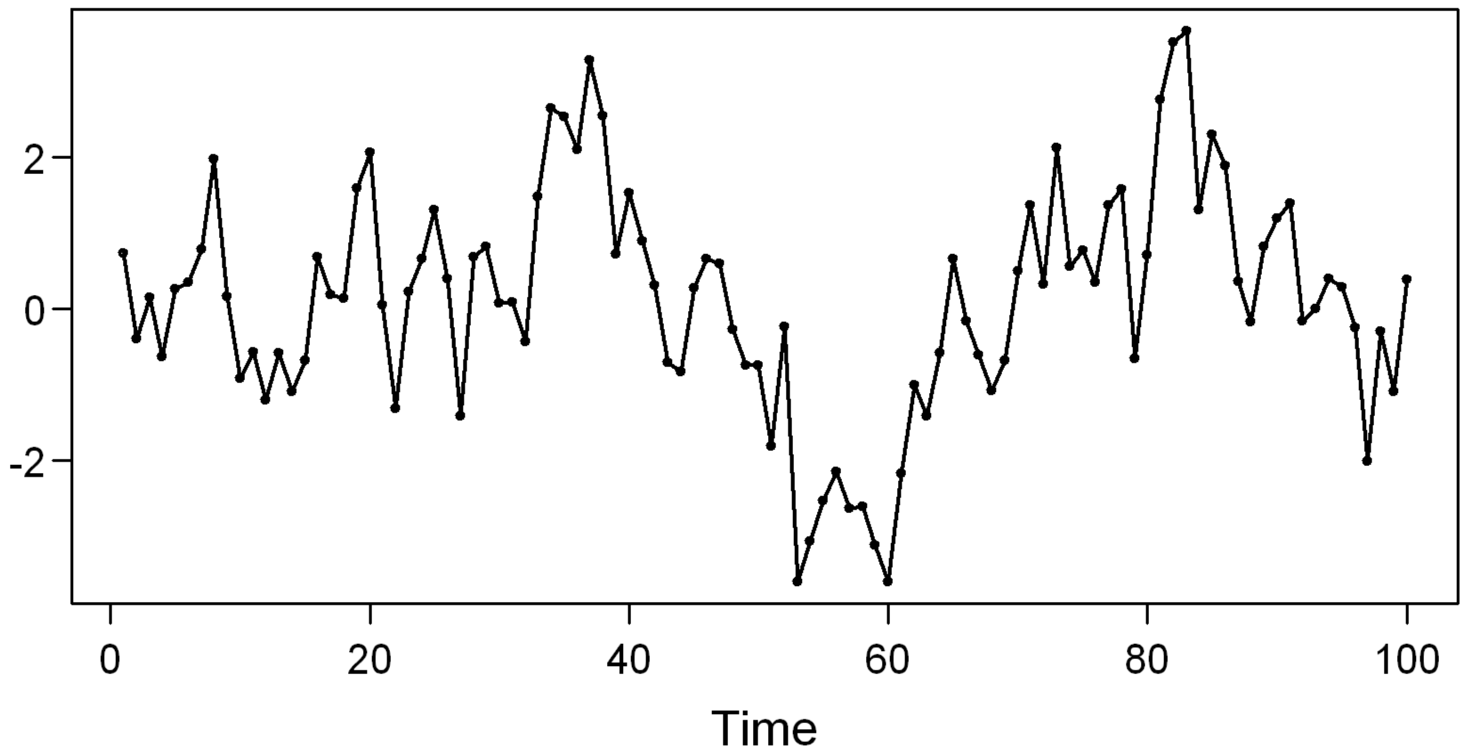
Monthly Price of West Texas Intermediate Crude Oil



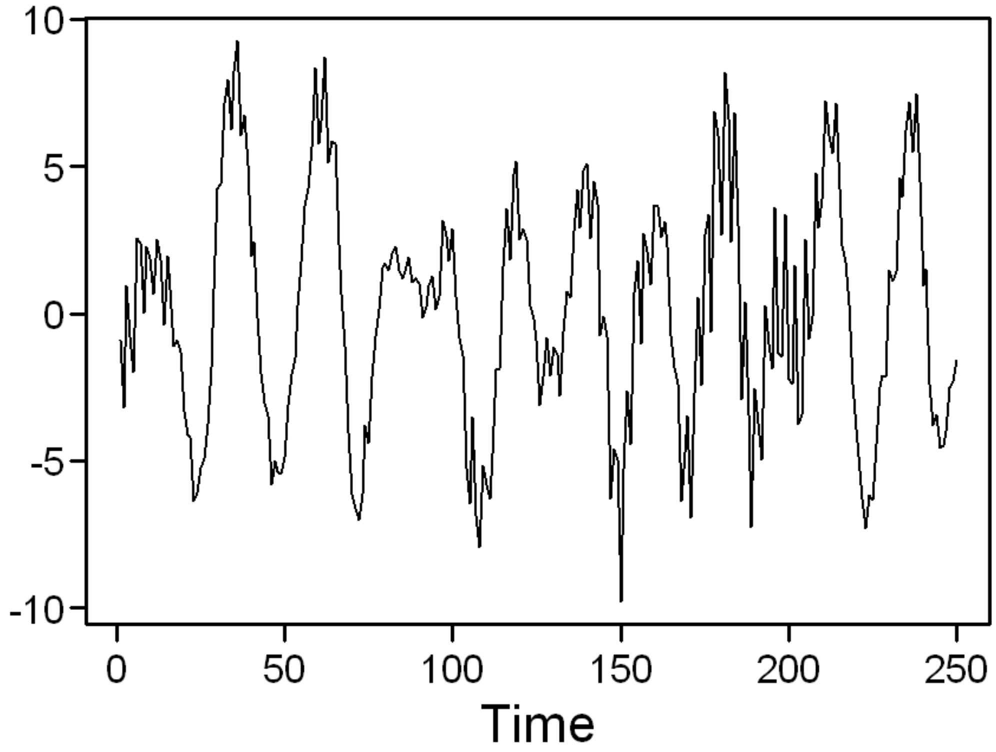
Average Monthly Temperature in Pennsylvania



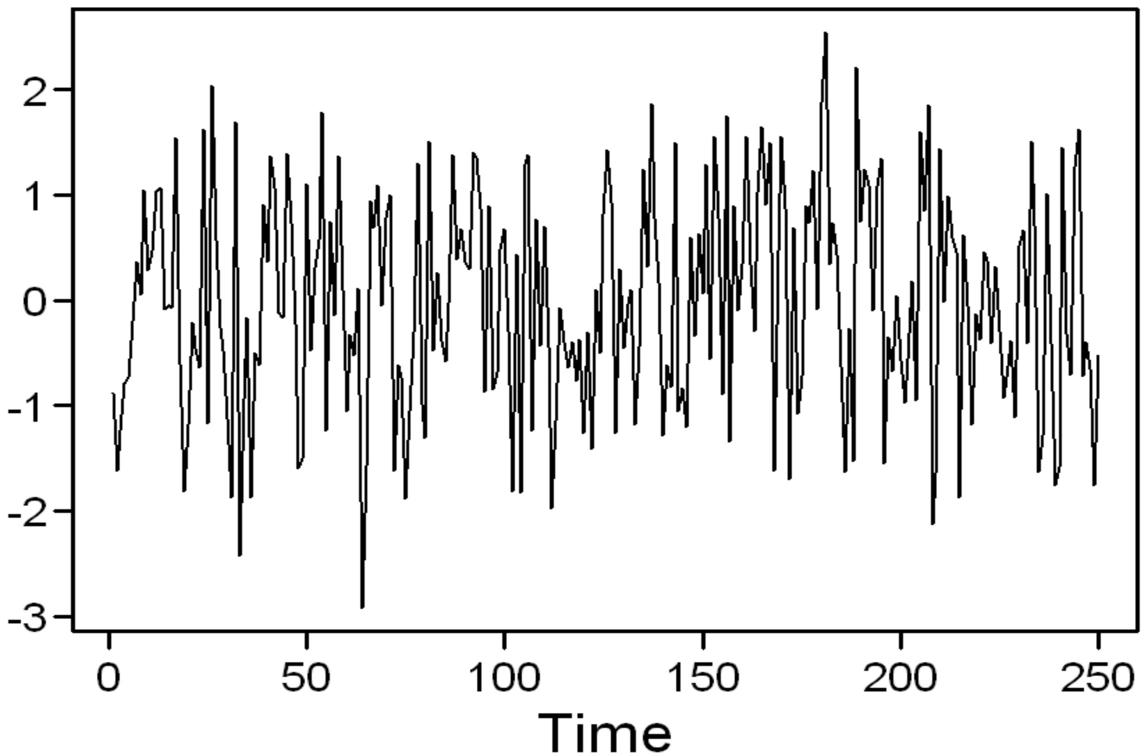
Simulated Data



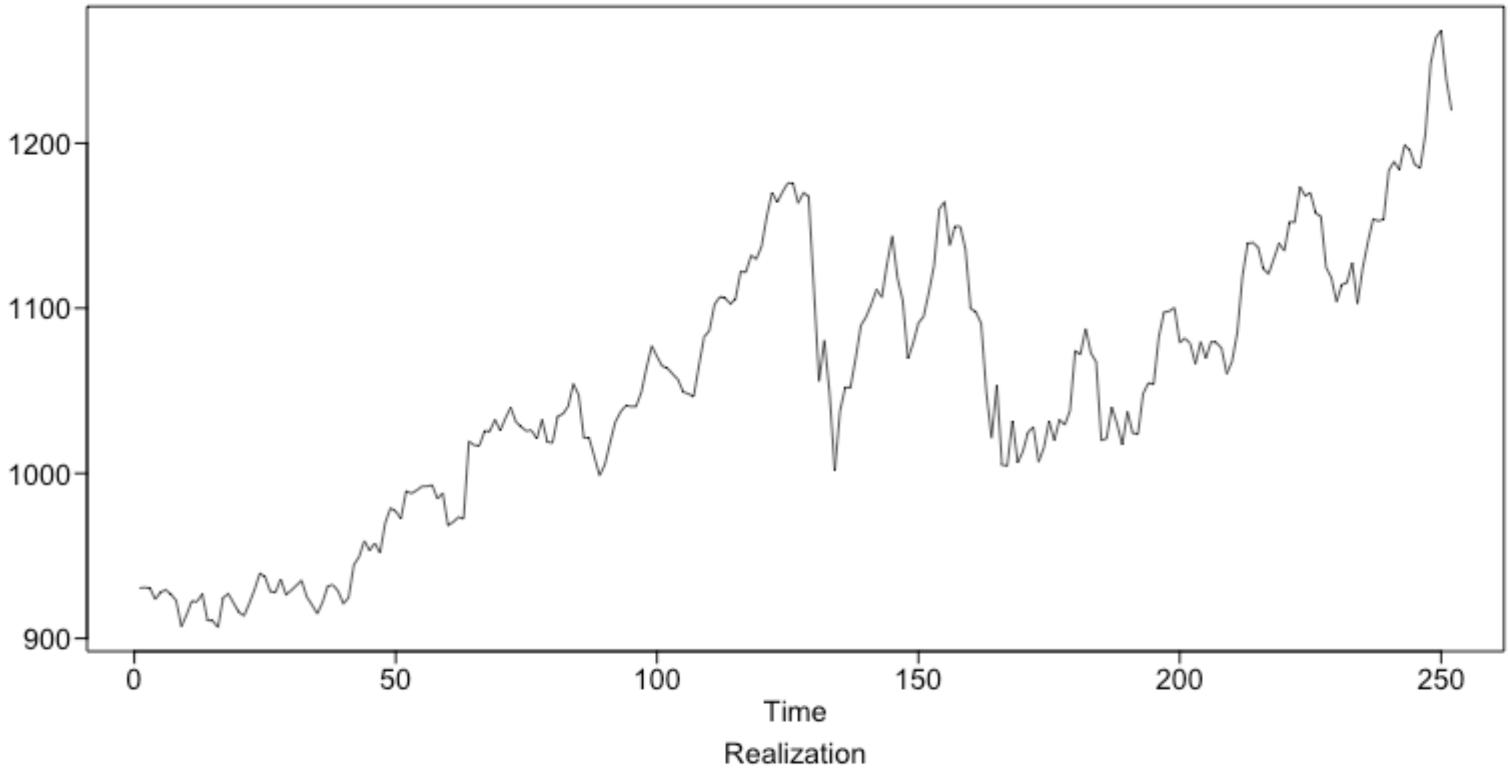
Simulated Data



Simulated Data

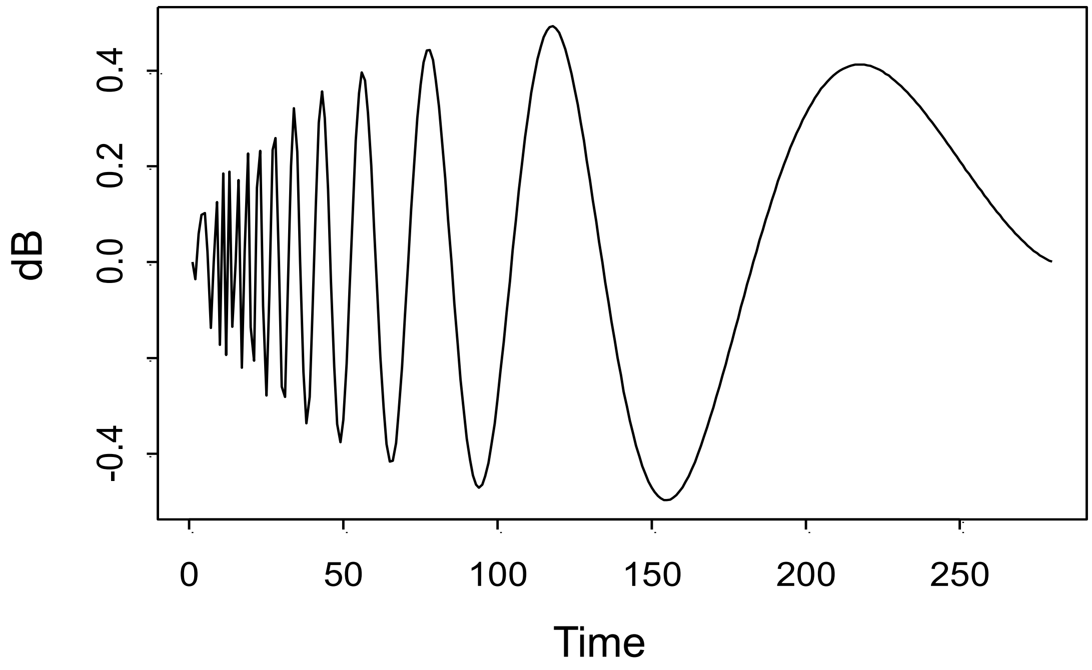


Google Stock Price

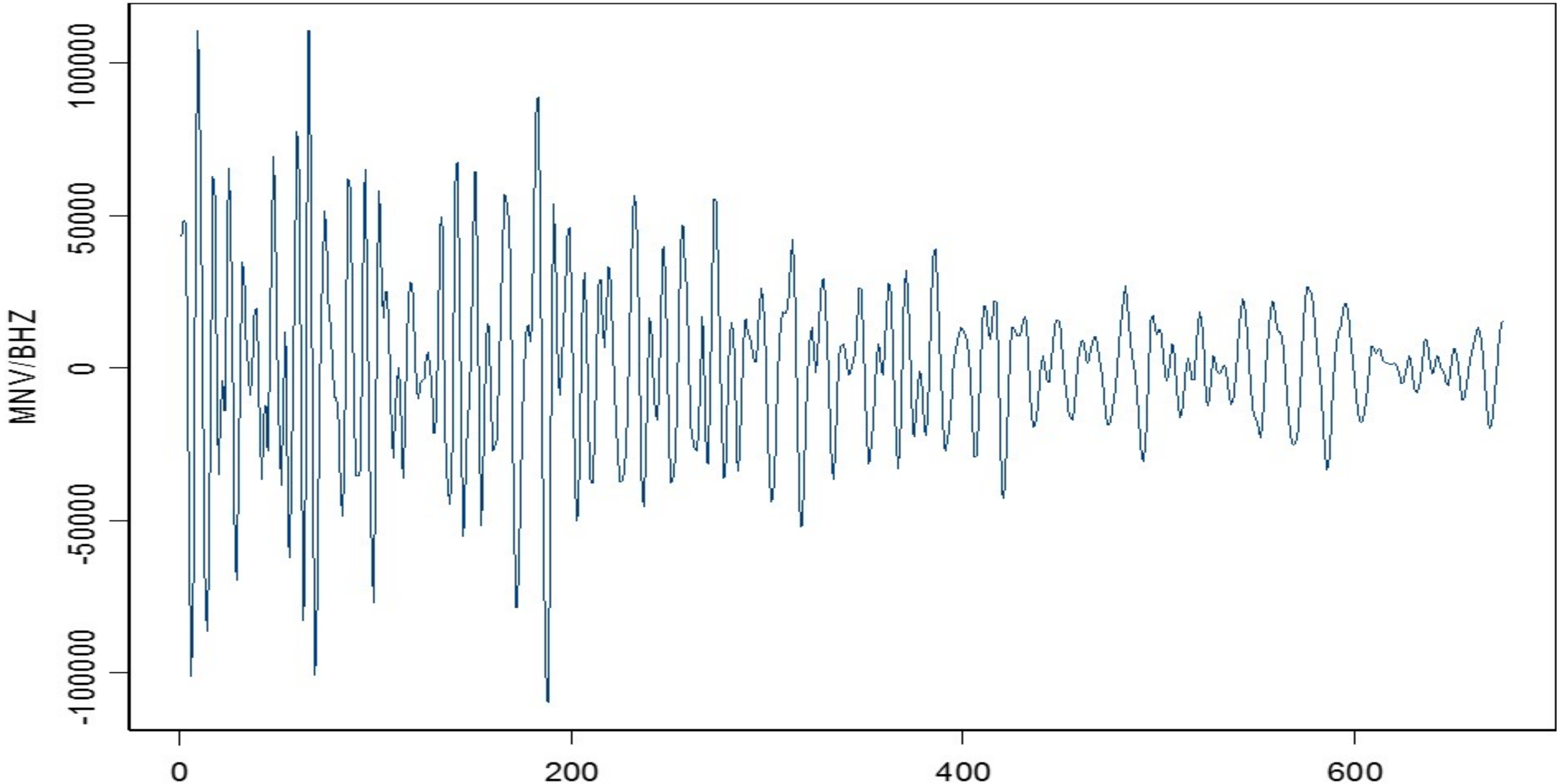


DataScience@SMU

Doppler Signal

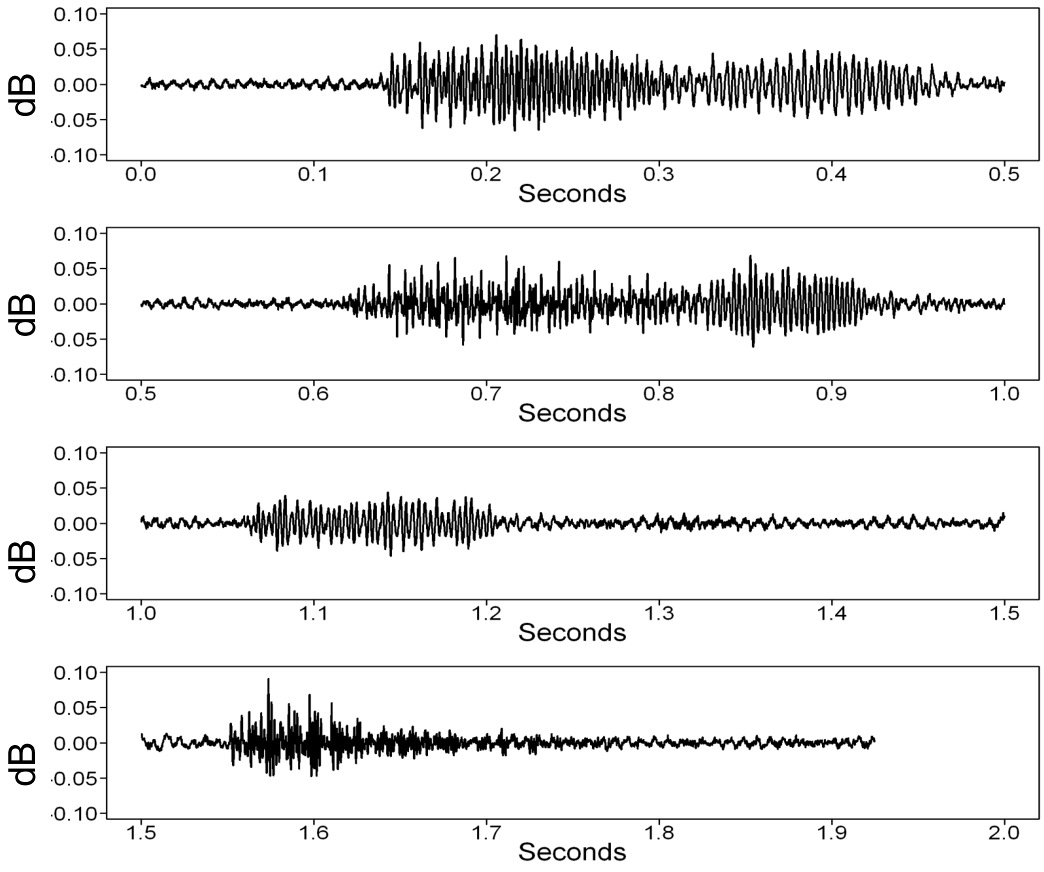


Seismic Wave of a Mining Explosion



Acoustical Signal of the Phrase “King Kong Eats Grass”

(Fan noise in background)

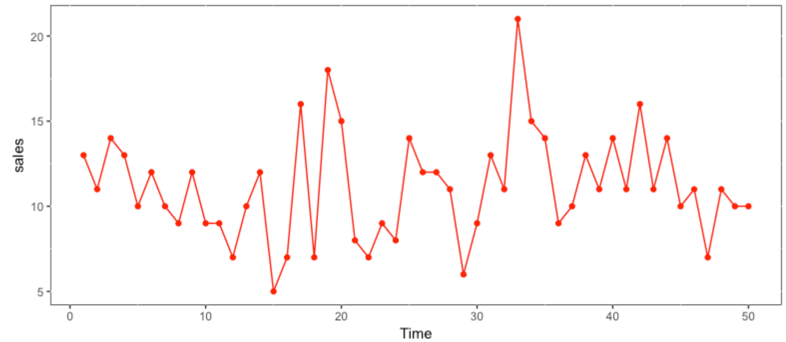


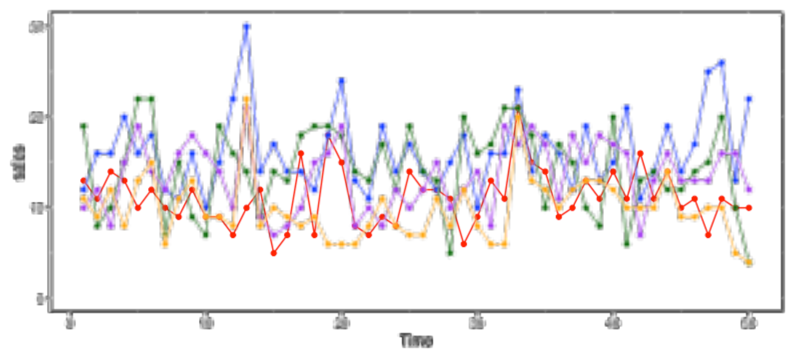
DataScience@SMU

Realization vs. Ensemble/Motivation

Realization

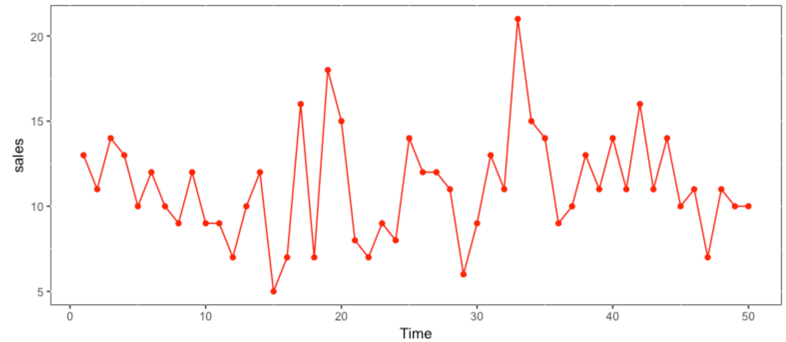
- A **realization** is a particular instance of time series.
- You can think of a **realization** as one of the infinitely many possible time series that we could have observed.
- The **realization** is the one we *actually* observed.
- Consider four more Walmarts that each sold dish soap on the same days.

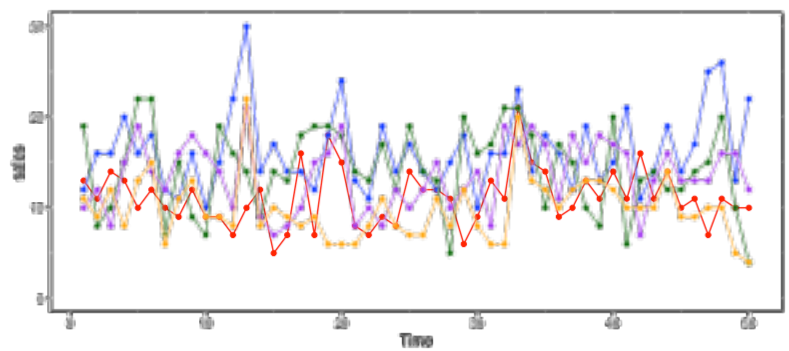


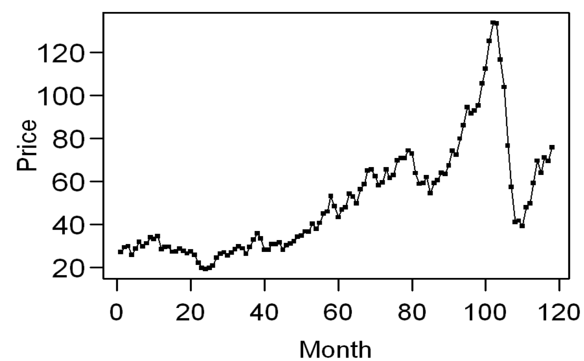


Realization

- A **realization** is a particular instance of time series.
- You can think of a **realization** as one of the infinitely many possible time series that we could have observed.
- The **realization** is the one we *actually* observed.
- Consider four more Walmarts that each sold dish soap on the same days.
- Sometimes it is impossible to obtain more than one realization. Take for instance West Texas crude oil prices.

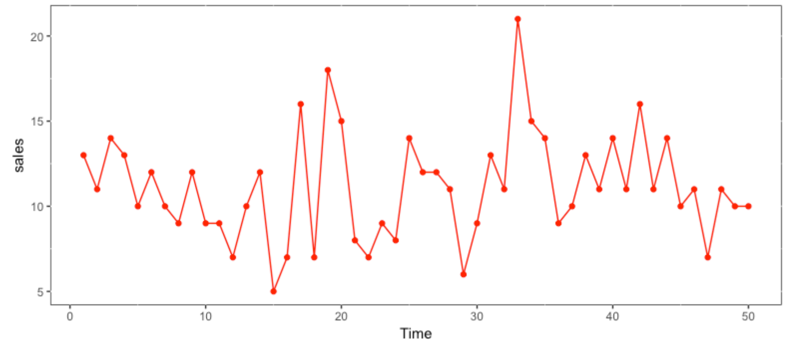


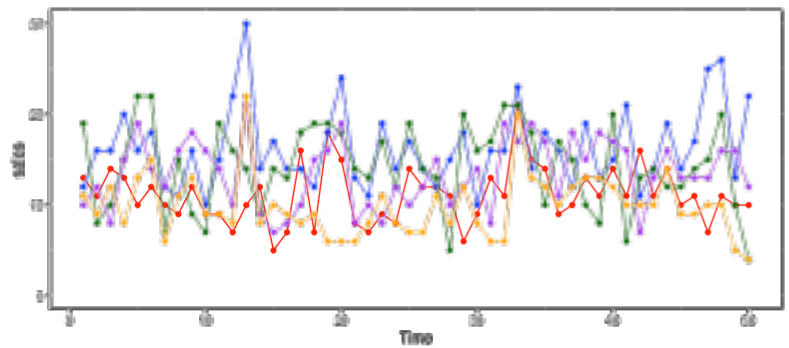


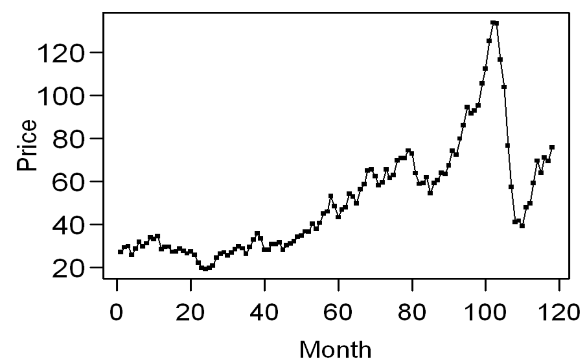


Realization

- A **realization** is a particular instance of time series.
- You can think of a **realization** as one of the infinitely many possible time series that we could have observed.
- The **realization** is the one we *actually* observed.
- Consider four more Walmarts that each sold dish soap on the same days.
- Sometimes it is impossible to obtain more than one realization. Take for instance West Texas crude oil prices.
- The **ensemble** is the totality of all possible realizations. (It is often impractical or impossible to view.) Ensemble equals population.







The Mean and Variance of X_t

We will be discussing (and estimating) the mean and variance of the random variable X_t.

Mean of X_t: $E[X_t] = \mu_t$

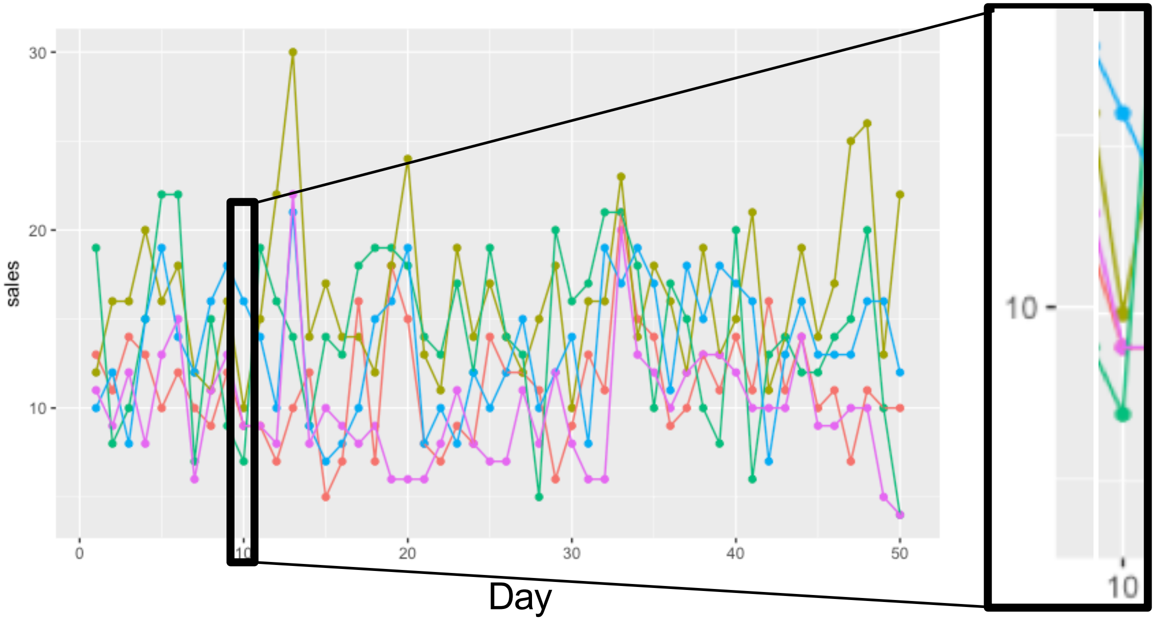
μ_t is the mean of all possible realizations of X_t ***for a fixed t.***

Variance of X_t: Var[X_t] = σ_t^2

σ_t^2 is the variance of all possible realizations of X_t ***for a fixed t.***

Example

We do not have the ensemble of all realizations from the Walmart example, but from the five we are given, let's estimate μ_t and σ_t^2 for t = 10.



$$\hat{\mu}_{10} = \frac{16 + 10 + 10 + 9 + 6}{5}$$

$$\hat{\mu}_{10} = 10.2$$

$$\hat{\sigma}_{10}^2 = 13.2$$

Student *t* (not time)

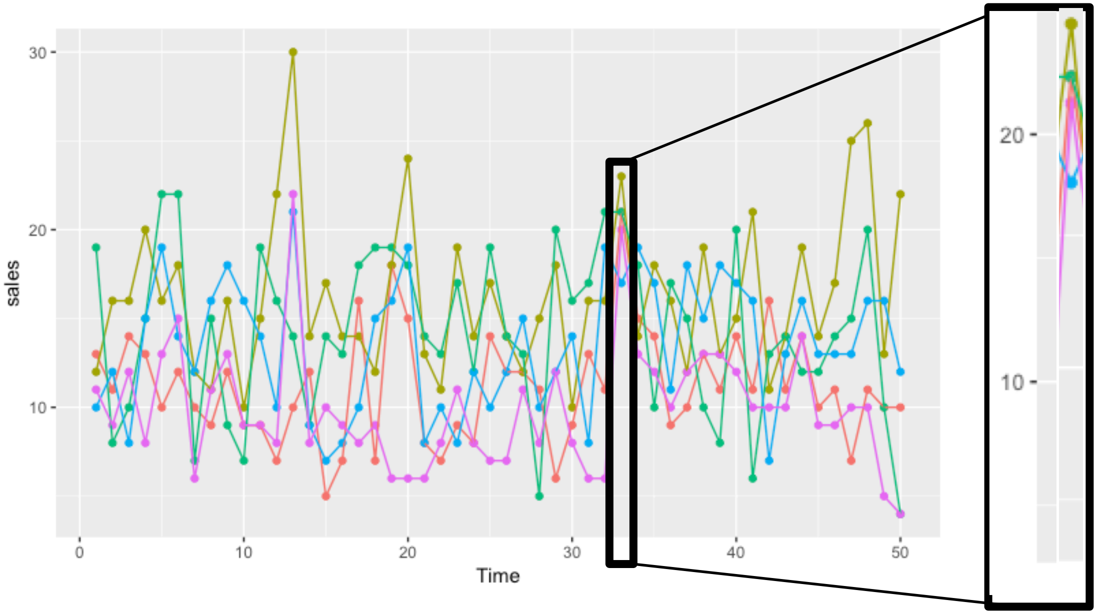
95% CI (assuming normality):

$$10.2 \pm t_{.975,4} * 3.63/\sqrt{5}$$

The means and variances could have been estimated in this fashion for all of the 50 possible values of *t* in this time series.

Example

We do not have the ensemble of all realizations from the Walmart example, but from the five we are given, let's estimate μ_t and σ_t^2 for *t* = 10.



$$\hat{\mu}_{33} = \frac{20 + 20 + 21 + 24 + 17}{5}$$

$$\hat{\mu}_{10} = 20.4$$

$$\hat{\sigma}_{10}^2 = 6.3$$

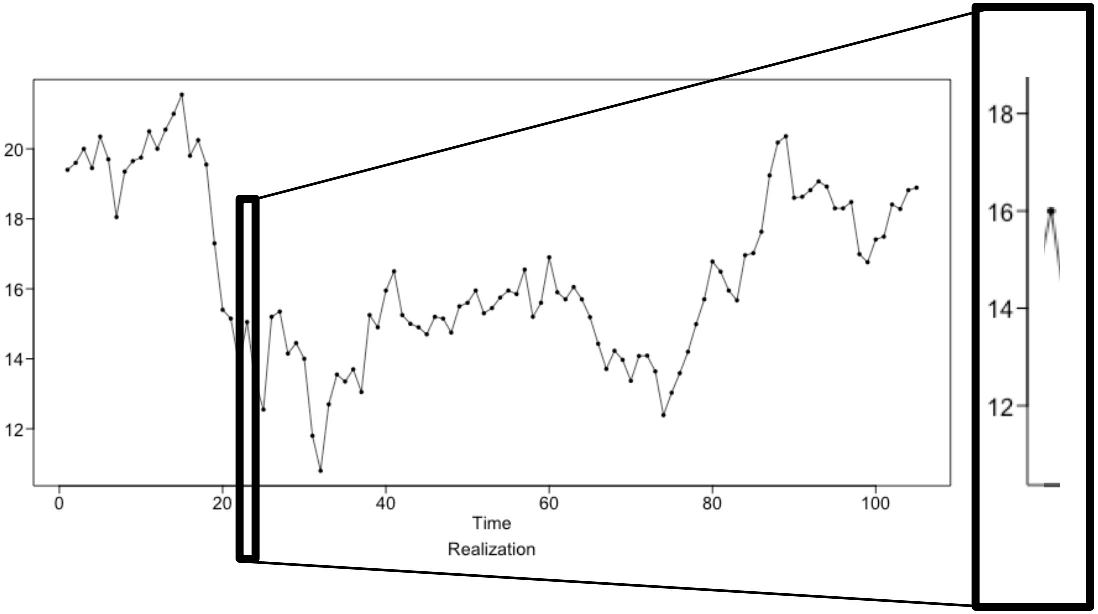
95% CI (assuming normality):

$$20.4 \pm t_{.975,4} * 2.51/\sqrt{5}$$

The means and variances could have been estimated in this fashion ***for each*** of the 50 possible values of *t* in this time series.

Example

In the time series setting, we typically only get one realization. Like the West Texas Intermediate crude data. We can't turn back time and see it all over again. There is only one reality/realization. Can we still estimate the mean and variance and construct a 95% confidence interval?



$$\hat{\mu}_{23} = \frac{16}{1} = 16$$

$$\hat{\mu}_{23} = 16$$

$$\hat{\sigma}^2_{23} = ?$$

95% CI (assuming normality):

$$16 \pm t_{.975,0} * ?$$

For each time point (*t*) there is only one observation, and thus, while we can theoretically get a trivial estimate of the mean, we cannot estimate the variance with only one point. **What can we do?**

DataScience@SMU

Expected Value

Expected Value

In short, the expected value of a random variable *X* **(*E[X]*) is the mean** or intuitively the long-run average of the event that variable represents.

For example:

Consider a variable *X* that represents the amount of money won in a particular casino game. *X* represents the amount won or lost (negative amount) and *P*(*X*) represents the probability of that event. We could organize this idea in a table:

X	*P(X)*
2	.1
0	.3
–1	.6

This indicates that there is a 10% chance that the player will win \$2, a 30% chance that they will break even (win \$0 but lose \$0), and a 60% chance that they will lose \$1.

Should casino goers play this game? If they played this game many times, what would be their average winnings or losings?

These questions can be addressed by looking at the expected win or loss (the expected value) each time they play.

$$\boldsymbol{Expected\ value\ of\ X = E[X] = \$2 * .1 + \$0 * .3 + (-\$1) * .6 = -\$0.4}$$

In this game, each time the player plays this game, he or she is expected to lose, on average, \$0.4, or 40 cents.

Spoiler alert, all games in Las Vegas have a negative expected value!

Expected Value

Recall, since *X* cannot take on any value between the ones listed above, we know that *X* is a **discrete** random variable.

X	*P(X)*
2	.1
0	.3
–1	.6

P(*X*) shows how the probabilities are distributed to the potential values of *X*. *P*(*X*) is known as a discrete probability distribution function (pdf).

$$\boldsymbol{Expected\ value\ of\ X = E[X]} = \$\mathbf{2} * .\mathbf{1} + \$\mathbf{0} * .\mathbf{3} + (-\$\mathbf{1}) * .\mathbf{6} = \mathbf{–\$0.4}$$

$$E[X] = \sum XP(X) = \mu$$

Formula for the expected value of a discrete random variable *X*

The expected value of *X* (*E*[*X*] is the population mean of the distribution)

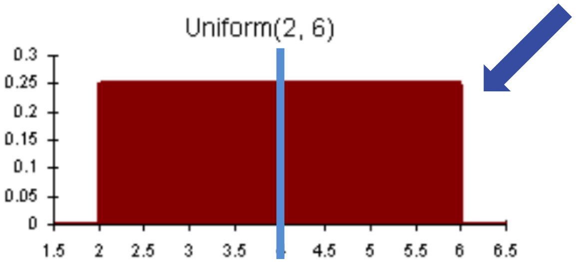
Expected Value

What about a continuous random variable *X*?

Continuous random variables also have probability distributions/density functions (*f*(*x*)).

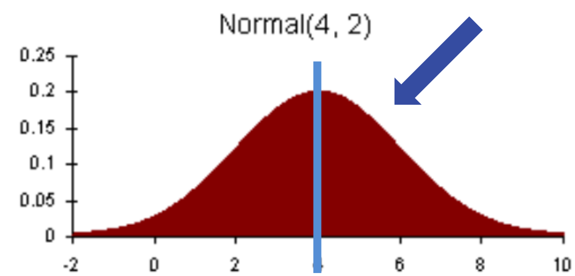
$$f(x) = \ .25\ for\ x \in [2,6]$$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{1}{2\sigma}(x-\mu)^2} for\ x \in [-\infty, \infty]$$



μ

$$E[X] = \int_a^b xf(x)dx$$



μ

$$E[X] = \int_2^6 xf(x)dx = \mu = 4$$

Main takeaway

$$\boxed{E[X] = \mu}$$

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx = \mu = 4$$

Expected Value

Properties

Let *a* and *b* be constants.

$$E[a] = a$$

$$E[aX] = aE[X] = a\mu$$

$$E[aX + b] = aE[X] + E[b] = aE[X] + b = a\mu\text{+b}$$

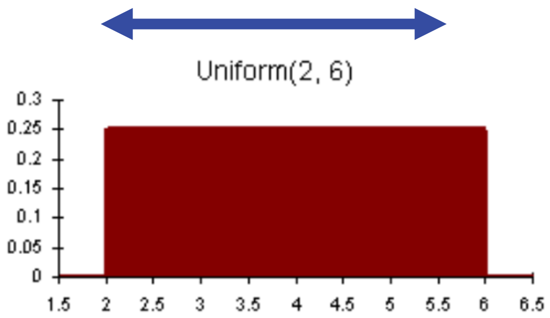
$Try\ one$: Find $E[\bar{x}]$ where $\bar{x} = \frac{\sum_{i=1}^{n} X_i}{n}$.

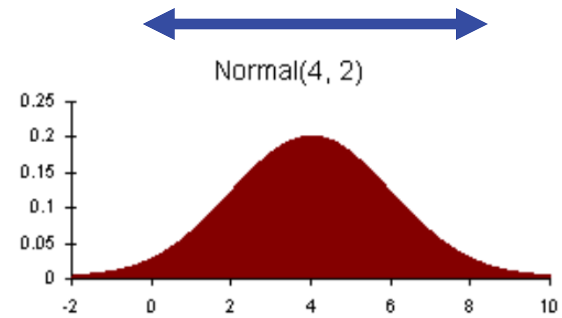
$$E[\bar{x}] = E\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}E\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n}n\mu = \mu$$

DataScience@SMU

Variance

Variance





$$\boxed{Var(X) = \mathrm{E}[(\mathrm{X} - \mu)^2]}$$

$$= \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$$

Main takeaways

$$= \sigma^2$$

$$\boxed{\widehat{Var(X)} = \sum (x_i - \bar{x})^2}$$

DataScience@SMU

Covariance

Covariance

Next, consider two random continuous variables *X* and *Y*.

We would like a measure of how much they move together.

That is, as *X* gets bigger, does *Y* tend to get smaller, larger, or is *Y* **independent** of *X*?

This can be measured with the **covariance** of *X* and *Y*: ***Cov(X,Y).***

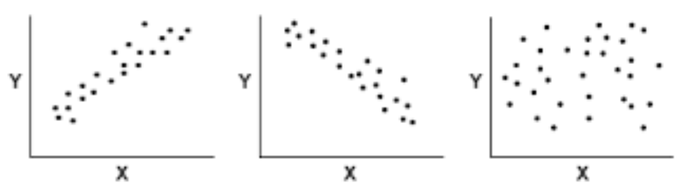
Estimate of the covariance of *X* and *Y* from the data

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

$$\widehat{Cov}(X,Y) = \sum (x_i - \bar{x})(y_i - \bar{y})$$

Known as the "sum of the cross products"

Main takeaway



***Cov(X,Y)* > 0** ***Cov(X,Y)* < 0** ***Cov(X,Y)* ≈ 0**

Use the above equations and the equations from the last slide to show these results: We will go over the answer in live session!

$$Cov(X,X) = Var(X)$$

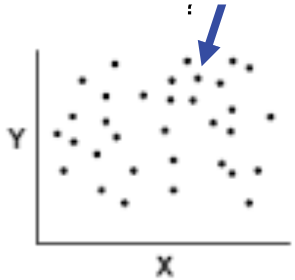
$$\widehat{Cov}(X,X) = \widehat{Var}(X)$$

DataScience@SMU

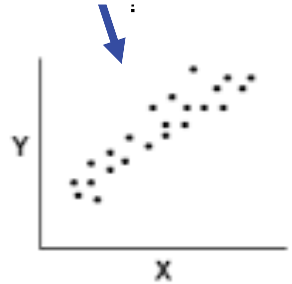
Correlation

Correlation

***Corr*(*X,Y*) = 0.005**

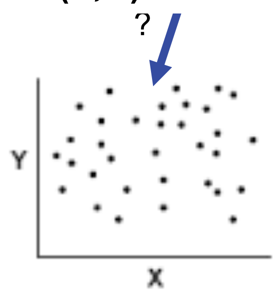


***Corr*(*X,Y*) = 0.921**

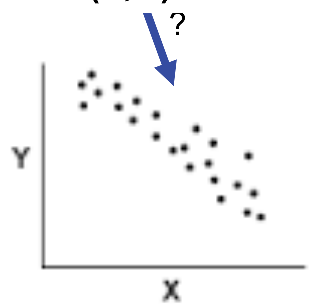


Depends* on the scale of *X* and *Y

***Corr*(*X,Y*) = 0.105**



***Corr*(*X,Y*) = –.8993**



The correlation between *X* and *Y* allows us to obtain a measure of dependence between *X* and *Y* that is scaled so that it must be between –1 and 1 inclusive.

In fact, the correlation, which you are very familiar with, is simply the scaled covariance; we simply divide the covariance by the standard deviation of both *X* and *Y* and we have the correlation.

$$Corr(X,Y) = \frac{Cov(X,Y)}{SD(X)SD(Y)}$$

Main takeaway

$$= \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

$$\widehat{Corr(X,Y)} = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{s_X s_Y}$$

DataScience@SMU

Condition 1

Mean Does Not Depend on Time

Stationary Series

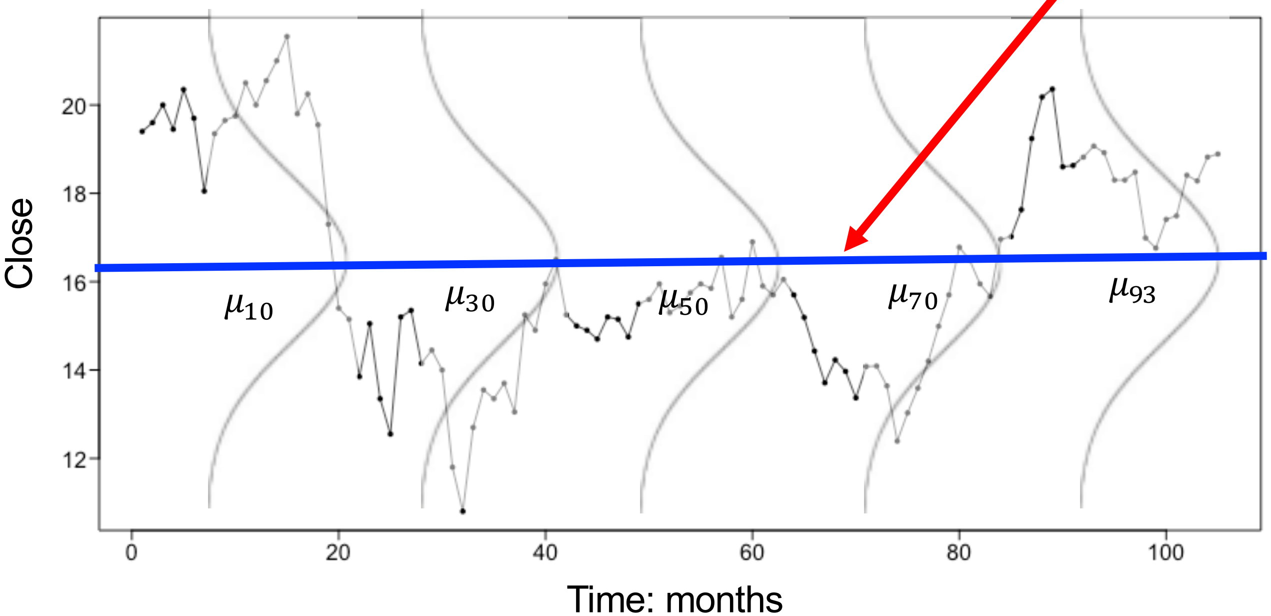
Three conditions:

1. Subpopulations of X_t have the same mean for each *t*. Restated, the mean does not depend on time (*t*).

$$E[X_t] = \mu$$

Note: no *t* subscript!

The population mean: $\mu = 17$



If we assume a constant mean across all *t* (horizontally), it is reasonable to use all the data to estimate that common mean.

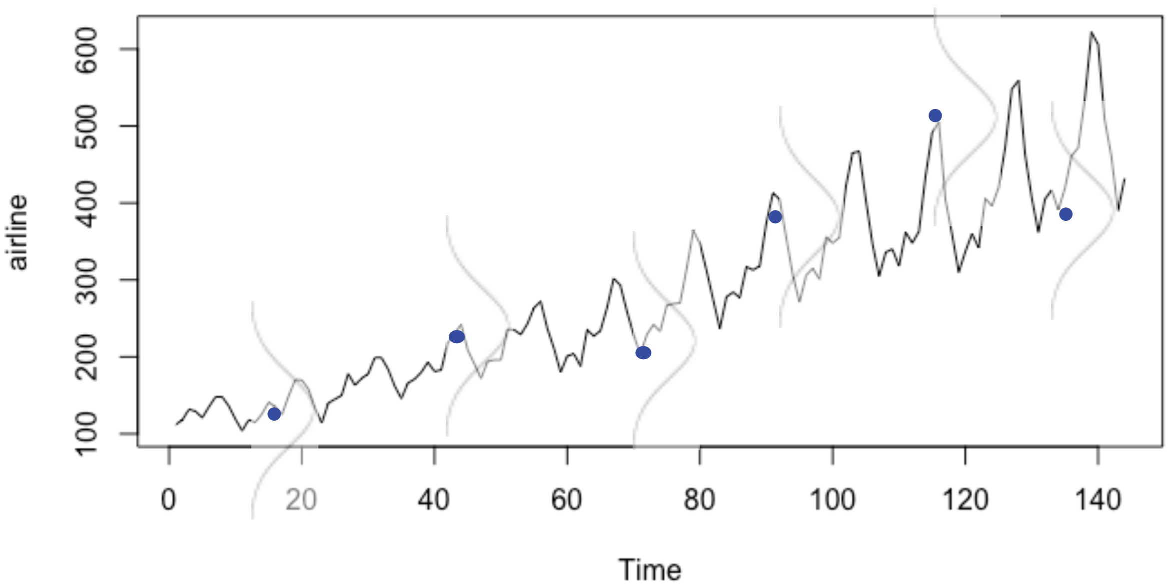
$$\hat{\mu}_{10} = \hat{\mu}_{82} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\bar{x} = 15.8$$

*Normal distributions are commonly assumed but not a requirement for stationarity.

$$\mu_{10} = \mu_{30} = \mu_{50} = \mu_{70} = \mu_{93} = \mu = 17$$

Example Where a Constant Mean for Each Time Point (t) Is Not a Reasonable Assumption



Stationary Series

Three conditions:

1. Subpopulations of X_t have the same mean for each *t*. Restated, the mean does not depend on time (*t*).

DataScience@SMU

Stationary Mean

Screencast for Stationary Mean

```
a1 = gen.sigplusnoise.wge(100,coef=c(5,0),freq = c(.1,0), psi = c(3,0),vara = 3, plot = FALSE)
b1 = gen.sigplusnoise.wge(100,coef=c(5,0),freq = c(.1,0), psi = c(3,0),vara = 3, plot = FALSE)
c1 = gen.sigplusnoise.wge(100,coef=c(5,0),freq = c(.1,0), psi = c(3,0),vara = 3, plot = FALSE)
d1 = gen.sigplusnoise.wge(100,coef=c(5,0),freq = c(.1,0), psi = c(3,0),vara = 3, plot = FALSE)
e1 = gen.sigplusnoise.wge(100,coef=c(5,0),freq = c(.1,0), psi = c(3,0),vara = 3, plot = FALSE)
plot(a1,type = "l")
lines(b1,col = "blue", type = "l")
lines(c1,col = "red", type = "l")
lines(d1,col = "green", type = "l")
lines(e1,col = "purple", type = "l")
```

```
a2 = gen.sigplusnoise.wge(100,coef=c(5,0),freq = c(.1,0), psi = c(runif(1,0,2*pi),0),vara = 3, plot = FALSE)
b2 = gen.sigplusnoise.wge(100,coef=c(5,0),freq = c(.1,0), psi = c(runif(1,0,2*pi),0),vara = 3, plot = FALSE)
c2= gen.sigplusnoise.wge(100,coef=c(5,0),freq = c(.1,0), psi = c(runif(1,0,2*pi),0),vara = 3, plot = FALSE)
d2 = gen.sigplusnoise.wge(100,coef=c(5,0),freq = c(.1,0), psi = c(runif(1,0,2*pi),0),vara = 3, plot = FALSE)
e2 = gen.sigplusnoise.wge(100,coef=c(5,0),freq = c(.1,0), psi = c(runif(1,0,2*pi),0),vara = 3, plot = FALSE)
plot(a2,type = "l")
lines(b2,col = "blue", type = "l")
lines(c2,col = "red", type = "l")
lines(d2,col = "green", type = "l")
lines(e2,col = "purple", type = "l")
```

```
par(mfrow = c(2,1))
plot(a1,type = "l")
lines(b1,col = "blue", type = "l")
lines(c1,col = "red", type = "l")
lines(d1,col = "green", type = "l")
lines(e1,col = "purple", type = "l")
plot(a2,type = "l")
lines(b2,col = "blue", type = "l")
lines(c2,col = "red", type = "l")
lines(d2,col = "green", type = "l")
lines(e2,col = "purple", type = "l")
```

DataScience@SMU

Condition 2

Variance Does Not Depend on Time

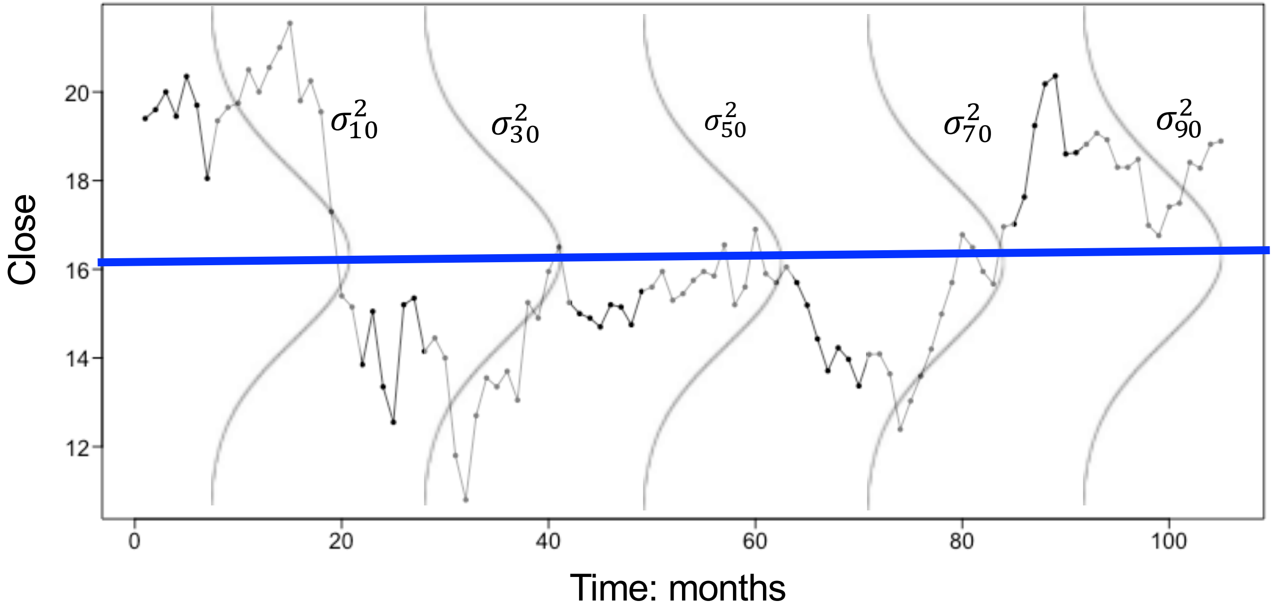
Stationary Series

Note: no *t* subscript!

Three conditions:

2. Subpopulations of *X* for a given time have a finite and constant variance for all *t*. Restated, the variance does not depend on time.

$$\text{Var}[X_t] = \sigma^2 < \infty$$



If we assume a constant variance for the distribution of *X* for each *t*, we can use all the data to estimate the common variance.

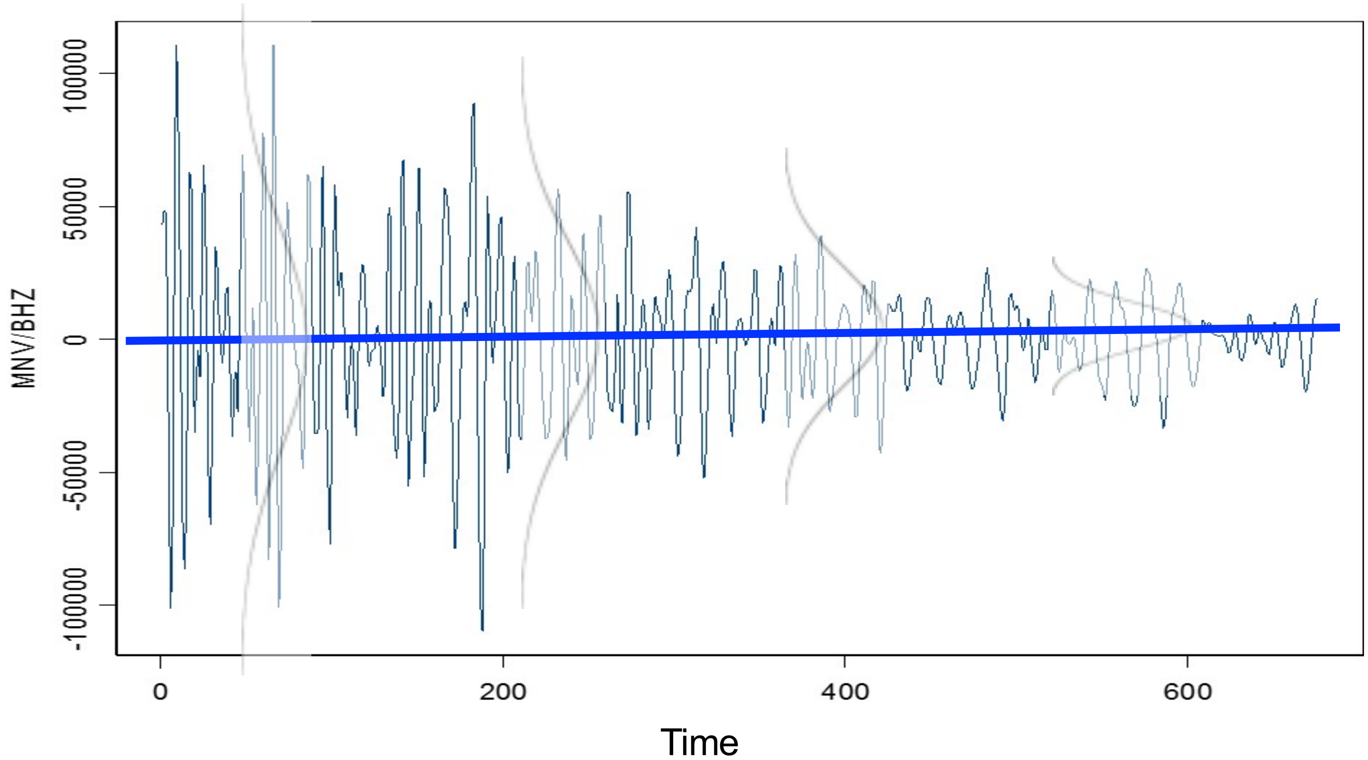
$$\hat{\sigma}^2_{10} = s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

*Normal distributions are commonly assumed but not a requirement for stationarity.

$$\sigma^2_{10} = \sigma^2_{30} = \sigma^2_{50} = \sigma^2_{70} = \sigma^2_{90} = \sigma^2$$

Seismic Data: Example Where a Constant Variance/SD for Each Time Point (t) Is Not a Reasonable Assumption*

*Assuming a constant mean of 0 for each time t



Stationary Series

Three conditions:

1. Subpopulations of X_t have the same mean for each t. Restated, **the mean does not depend on time (*t*).**
2. Subpopulations of X for a given time have a finite and constant variance for all t. Restated, **the variance does not depend on time.**

DataScience@SMU

Condition 3

Weak Stationary Process: Introduction

Weak Stationary Process

A weak stationary process is one that can be thought of as being in a state of "equilibrium."

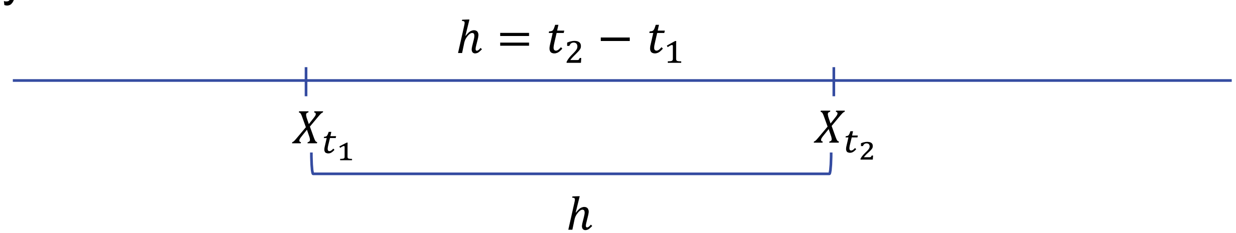
1. Subpopulations of X_t for a given time have constant mean for all *t*.

$$E[X_t] = \mu$$

2. Subpopulations of X_t for a given time have a constant variance for all *t*.

$$\text{Var}[X_t] = \sigma^2$$

3. The correlation of X_{t_1} and X_{t_2} depends only on $t_2 - t_1$. That is, the correlation between data points depends only on how far apart they are in time, not where they are in time.



$$Notation: Cor(X_t, X_{t+h}) = \rho_h$$

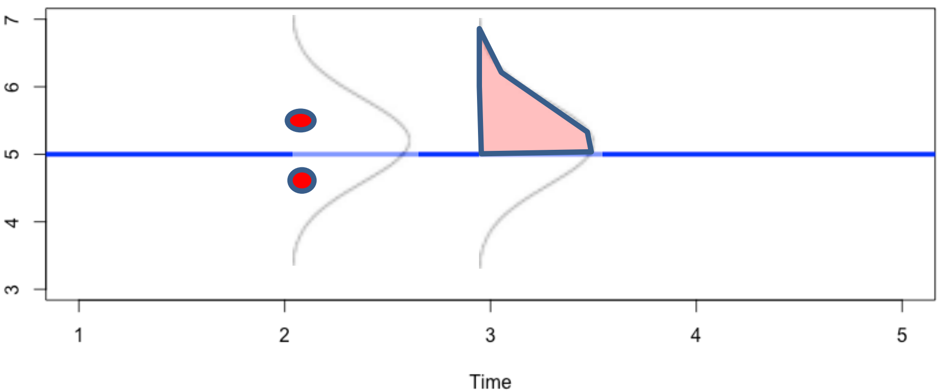
Before we discuss condition 3, let's review independence and correlation.

DataScience@SMU

Weak Stationary Process

Independence

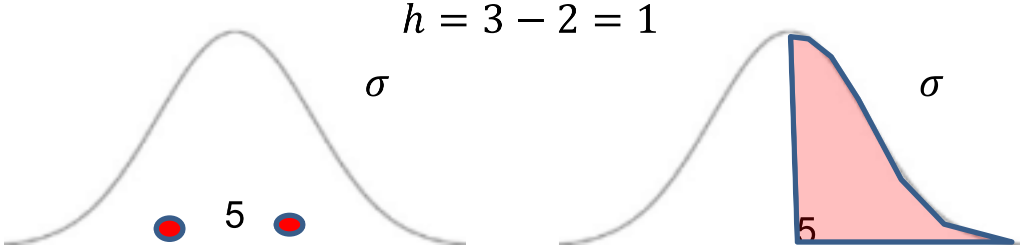
Independence



Time = 2

Time = 3

$$h = 3 - 2 = 1$$



Distribution of X_t when Time = 2

Distribution of X_t when Time = 3

Written definition:(time series):

Two events are said to be independent if the probability that one event occurs in no way depends on the occurrence of the other event.

independent in a time series if the probability that an event at time t occurs in no way depends on the occurrence of any event in the past.

Mathematical definition:

Independence: P($X_{t+1}|X_t$) = P(X_{t+1})

Example:

$P(X_3 > 5 \mid X_2 = 4.5) = P(X_3 > 5) = 0.5$

$P(X_3 > 5 \mid X_2 = 5.5) = P(X_3 > 5) = 0.5$

DataScience@SMU

Independence in tswge

Independent Events Screen Cast

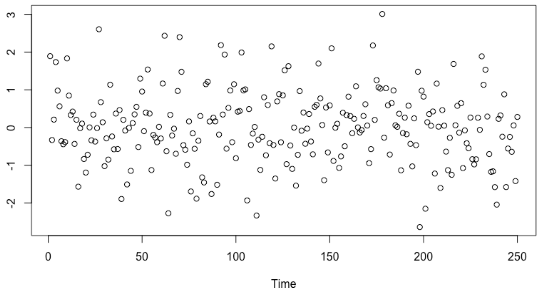
Assume that adjacent events are independent in time. That is, assume that the value of (X_1, … X_t) tells us no information about the value of X_{t+1}.

Mathematically, assume that $P(X_{t+1} \mid X_{1, \ldots} X_t) = P(X_{t+1})$.

In the time series setting, we call this "white noise." The notation is a_t.

For this course, $a_t \sim N(0, \sigma_a^2\ )\ such\ that\ \sigma_a^2 < \infty$.

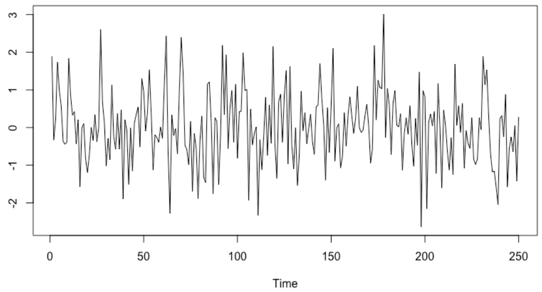
Let's use *R* and tswge to simulate this scenario, to simulate "white noise."



y = gen.arma.wge(*n* = 250)

Time = seq(1,250,length.out = 250)

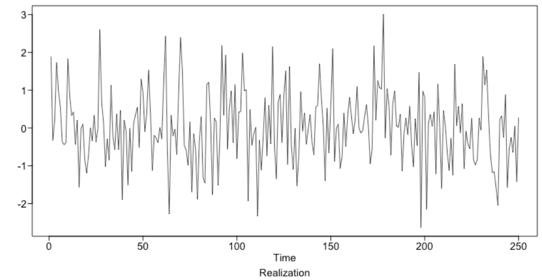
plot(Time,*y*)



y = gen.arma.wge(*n* = 250)

Time = seq(1,250,length.out = 250)

plot(Time,*y*,type = "l")

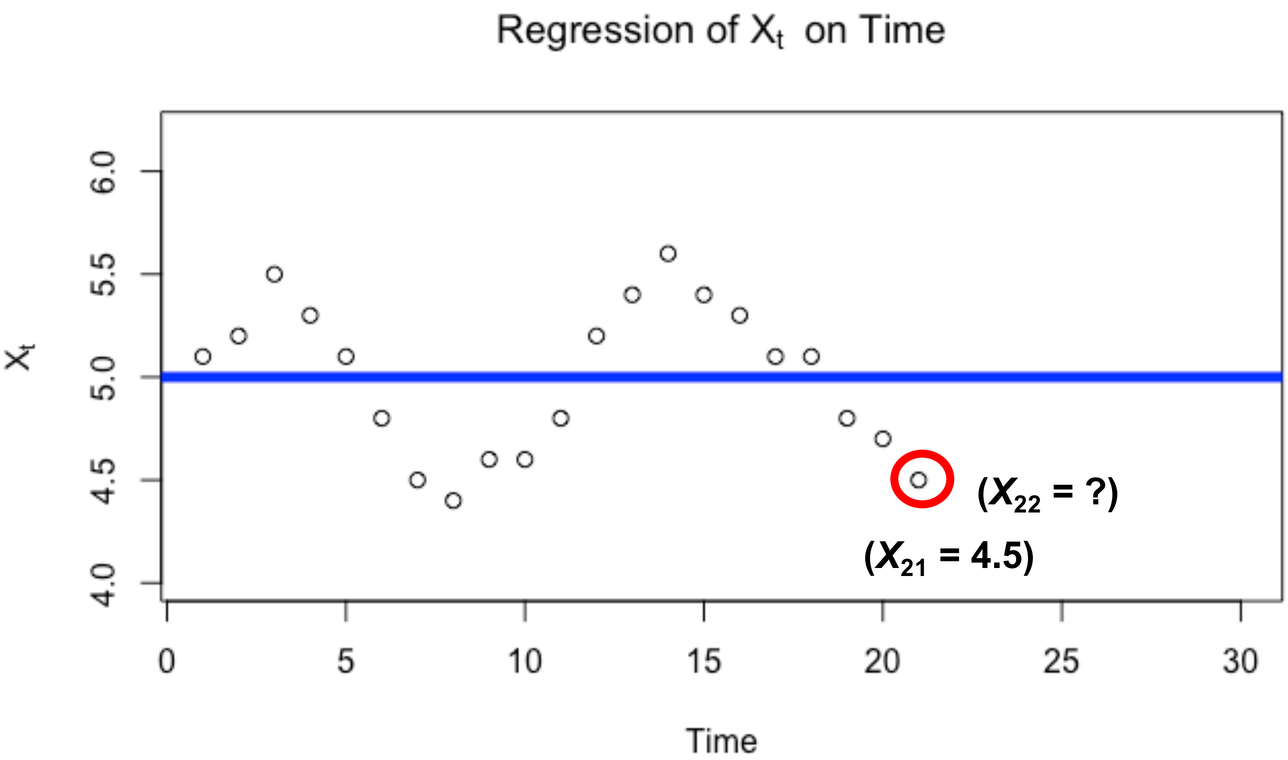


plotts.wge(*y*) #from tswge package

DataScience@SMU

Serial Dependence

Dependence: Serial Correlation



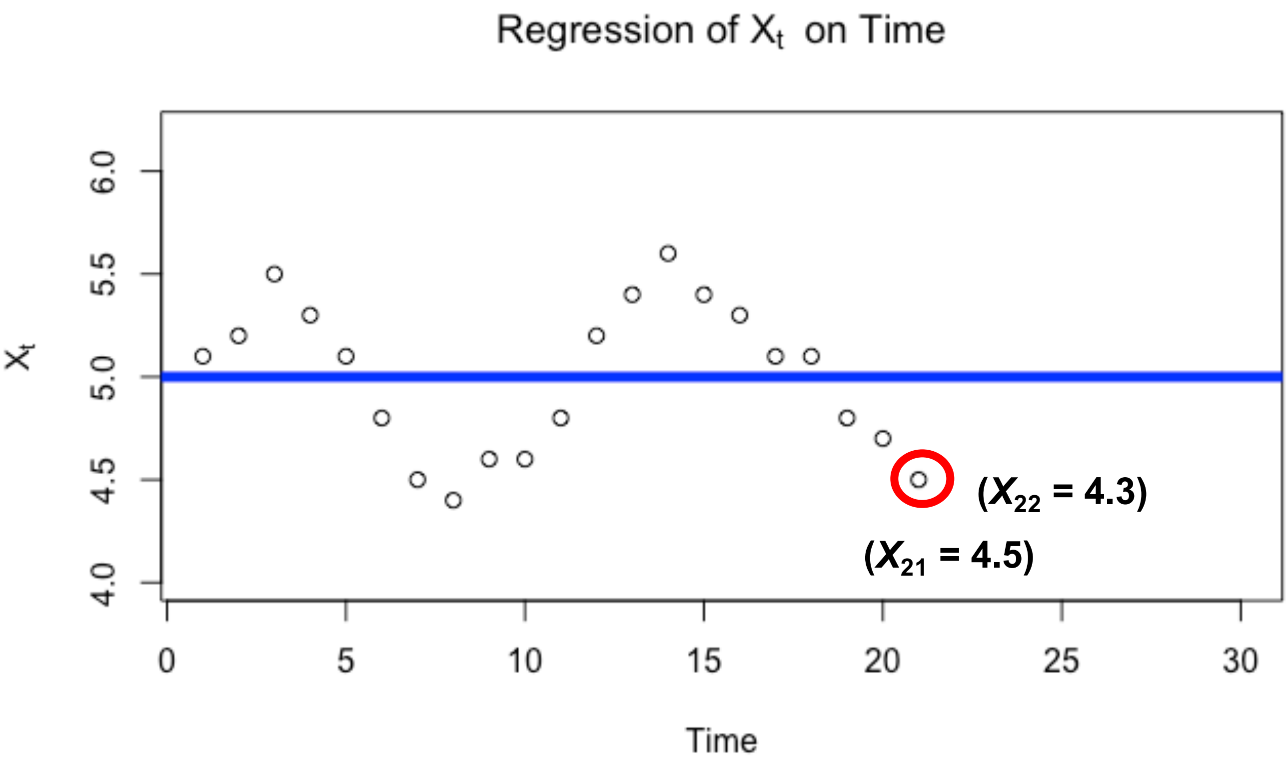
~~Example 1 (independent):~~

~~$P(X_{22} > 5 \mid X_{21} = 4.5) = P(X_{22} > 5) = .5$~~

Example 1 (dependent):

$P(X_{22} > 5 \mid X_{21} = 4.5) < P(X_{22} > 5) = .5$

Dependence: Serial Correlation



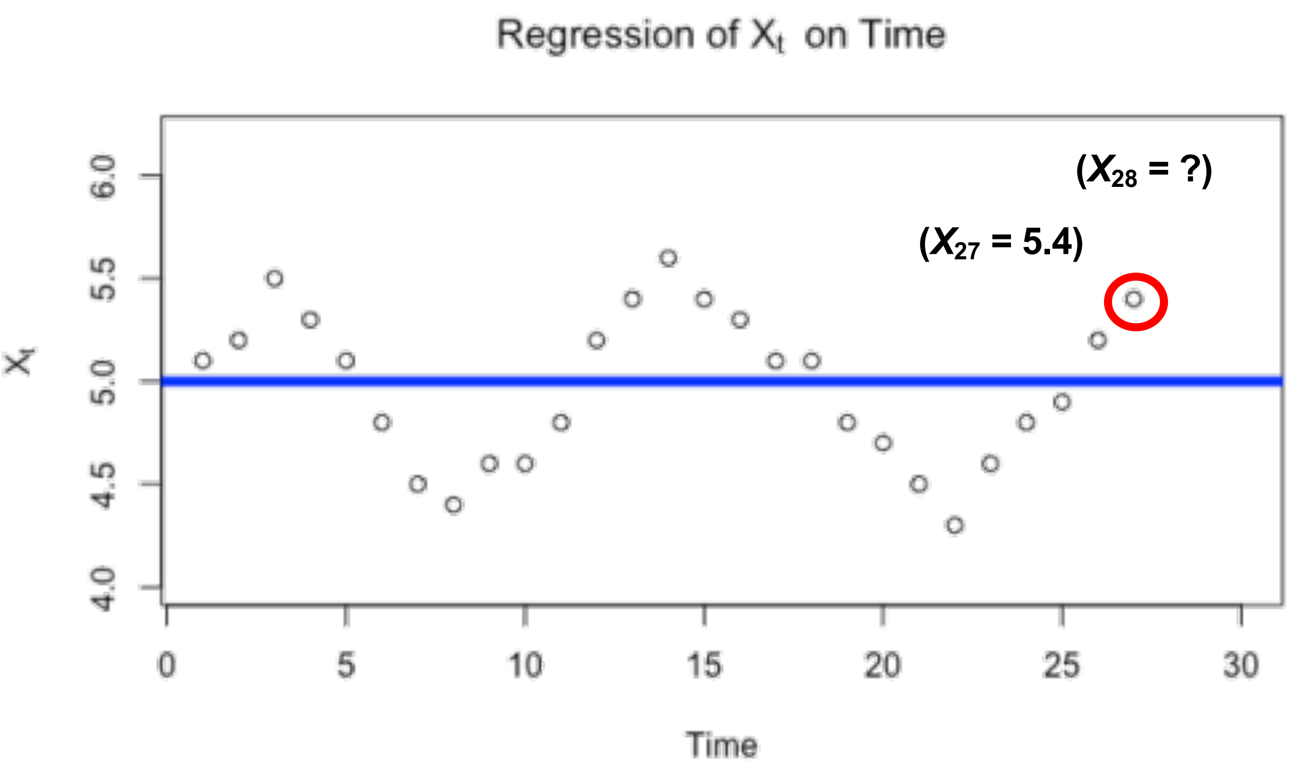
Example 1 (independent):

$P(X_{22} > 5 \mid X_{21} = 4.5) = P(X_{22} > 5) = .5$

Example 1 (dependent):

$P(X_{22} > 5 \mid X_{21} = 4.5) < P(X_{22} > 5) = .5$

Dependence: Serial Correlation



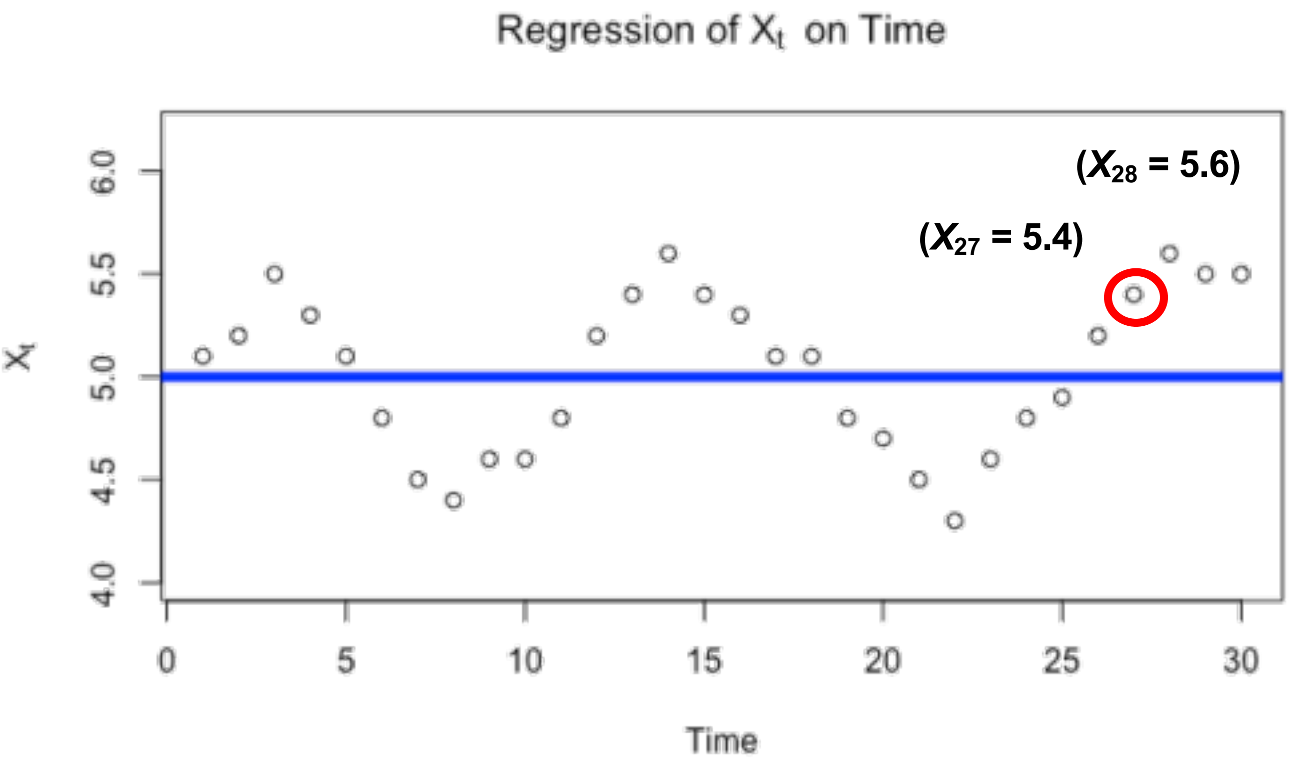
~~Example 1 (independent):~~

$\cancel{P(X_{28} > 5 \mid X_{27} = 5.4) = P(X_{28} > 5) = .5}$

Example 2 (dependent):

$P(X_{28} > 5 \mid X_{27} = 5.4) > P(X_{28} > 5) = .5$

Dependence: Serial Correlation



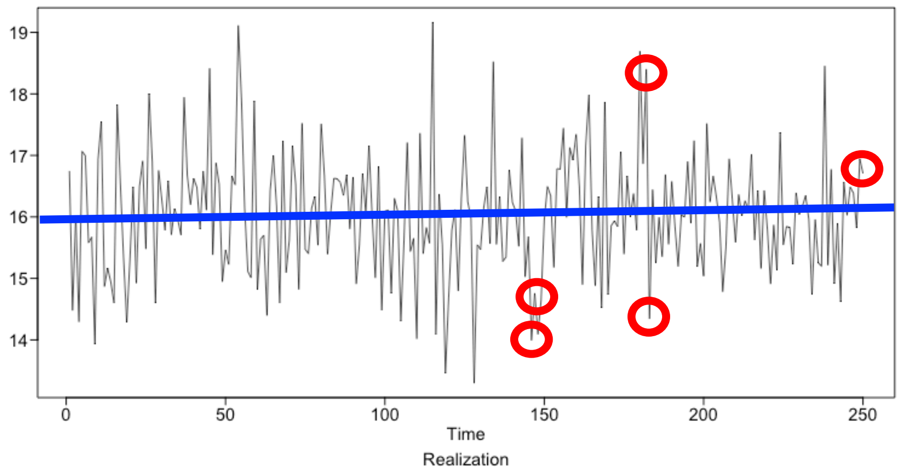
~~Example 1 (independent):~~

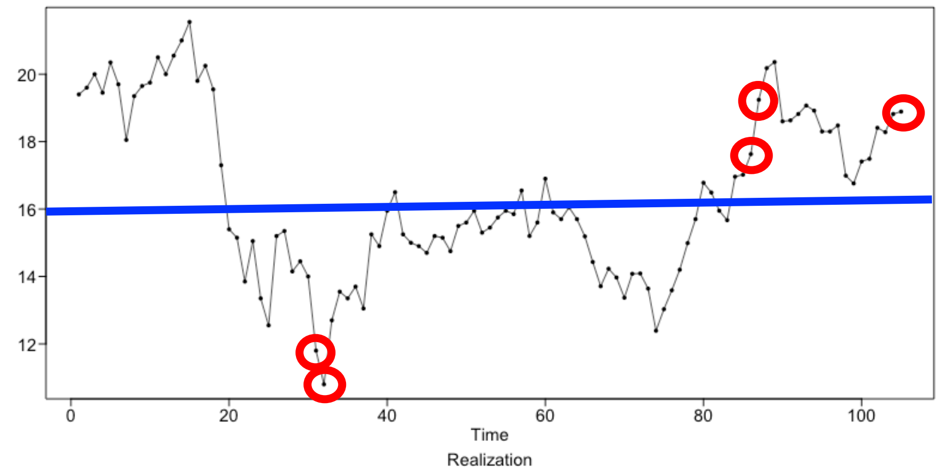
~~$P(X_{28} > 5 \mid X_{27} = 5.4) = P(X_{28} > 5) = .5$~~

Example 2 (dependent):

$P(X_{28} > 5 \mid X_{27} = 5.4) > P(X_{28} > 5) = .5$

Which Is Stronger Evidence of Dependence for X_t to X_{t+1}? Hint: What Will Be the Value of the Next Observation?





$P(X_{251} > 16 \mid X_{250} = 16.8) = P(X_{251} > 16) = .5$

$P(X_{101} > 16 \mid X_{100} = 19) > P(X_{101} > 16) = .5$

(4 min):
Weak Stationary Process
Slide Deck 3

Independence and Correlation

Note: 9/17/2018 These were hidden and not included since they are covered in the next section. We may decide to reinclude them based on how they flow when the lecture is practiced.

Independence and Correlation

- **Theorem: If two events are independent, their correlation is 0.**
  Recall that ρ is the population linear correlation coefficient.
  That is, if X_t and X_{t+k} are independent, $\rho_{X_t, X_{t+k}} = \rho_k = 0$.

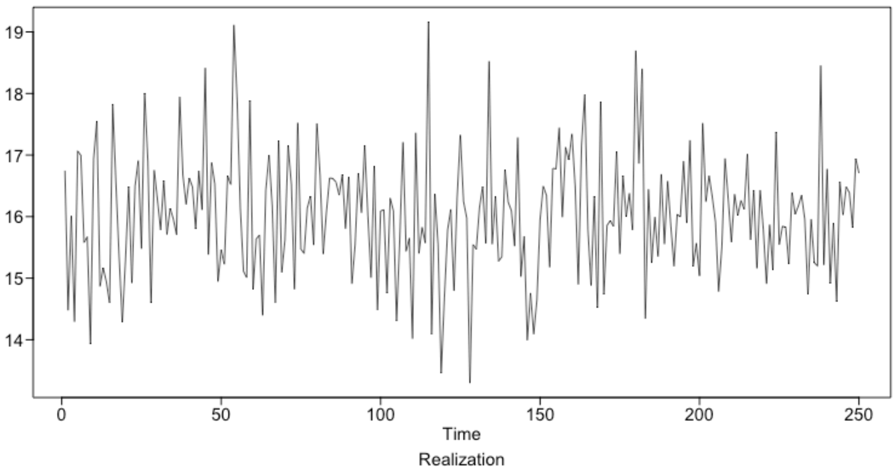
- **Corollary: If the correlation between two variables is nonzero, they are not independent.**
  That is, if $\rho_{X_t, X_{t+k}} = \rho_k \neq 0$, then X_t and X_{t+k} are not independent.

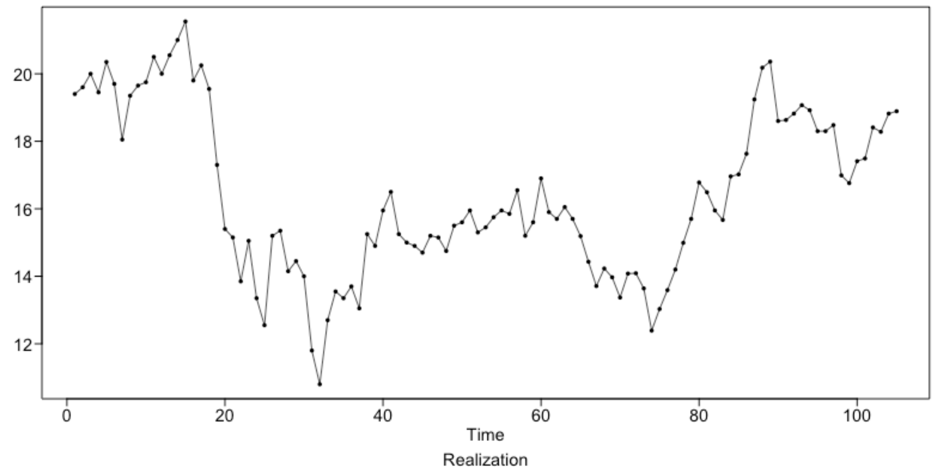
BLT 3 Concept Check (1 min):
Weak Stationary Process
Slide Deck 3

Independence and Correlation

Concept Check

Which is stronger evidence of $\rho_1 > 0$?
That is, which is stronger evidence that X_t and X_{t+1} are positively correlated?





$\rho_1 = 0$

$\rho_1 > 0$

DataScience@SMU

Autocorrelation

Independence and Correlation

- **Theorem: If two events are independent, their correlation is 0.**

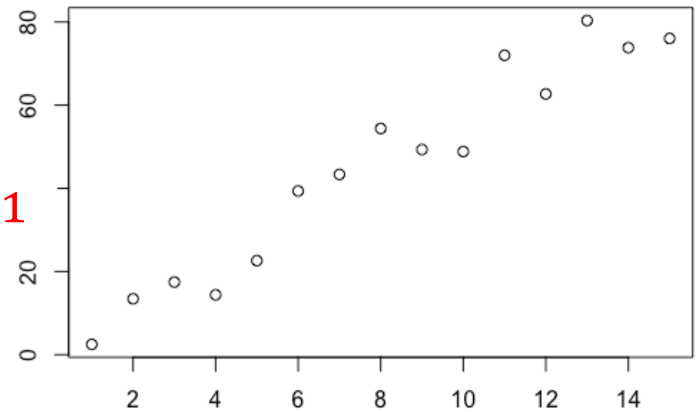
  Recall that ρ is the population linear correlation coefficient.

  That is, if X_t and X_{t+k} are independent, $\rho_{X_t,X_{t+k}} = 0$.

- **Corollary: If the correlation between two variables is nonzero, they are not independent.**

  That is, if $\rho_{X_t,X_{t+k}} \neq 0$, then X_t and X_{t+k} are not independent.

Recall: Correlation



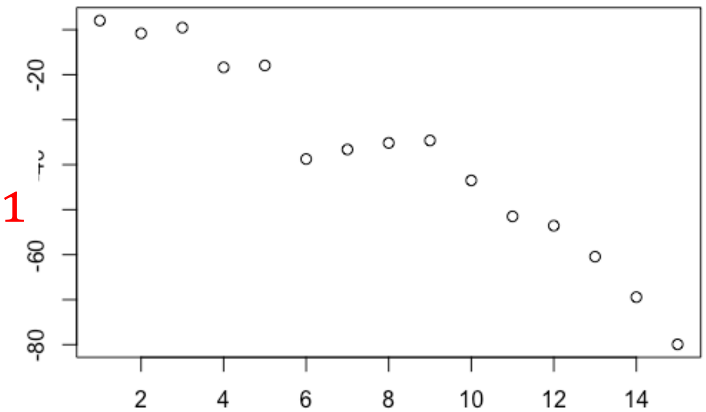
X_{t+1}

X_t

r = .9686

Lag (1)

Autocorrelation

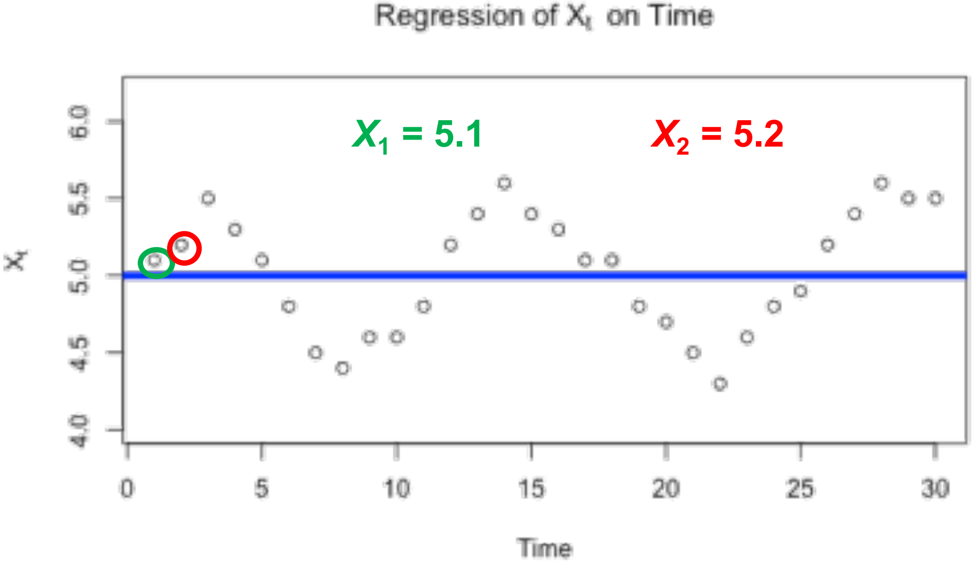


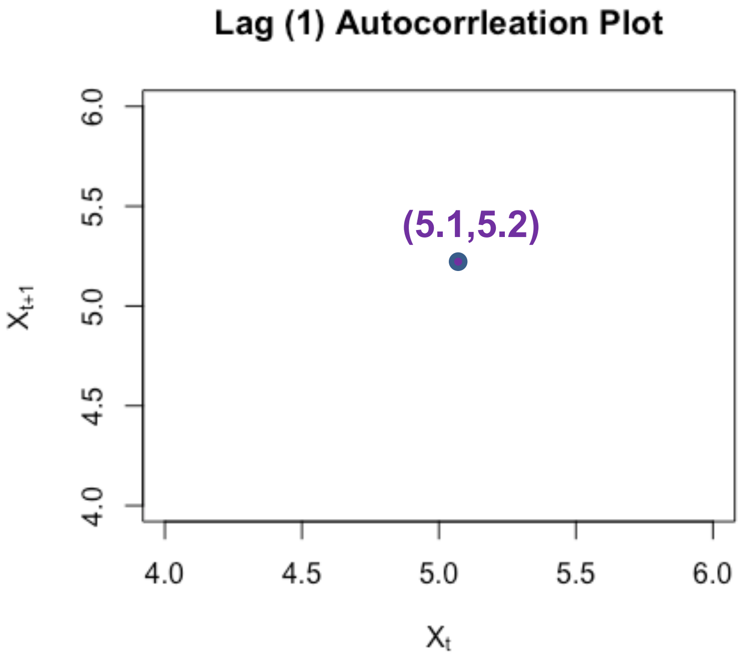
X_{t+1}

X_t

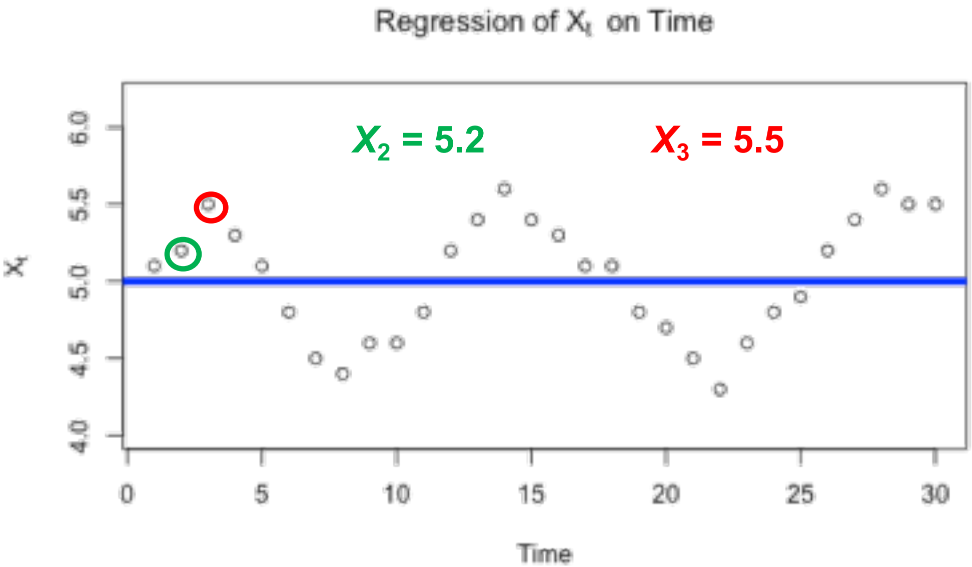
r = –.9730

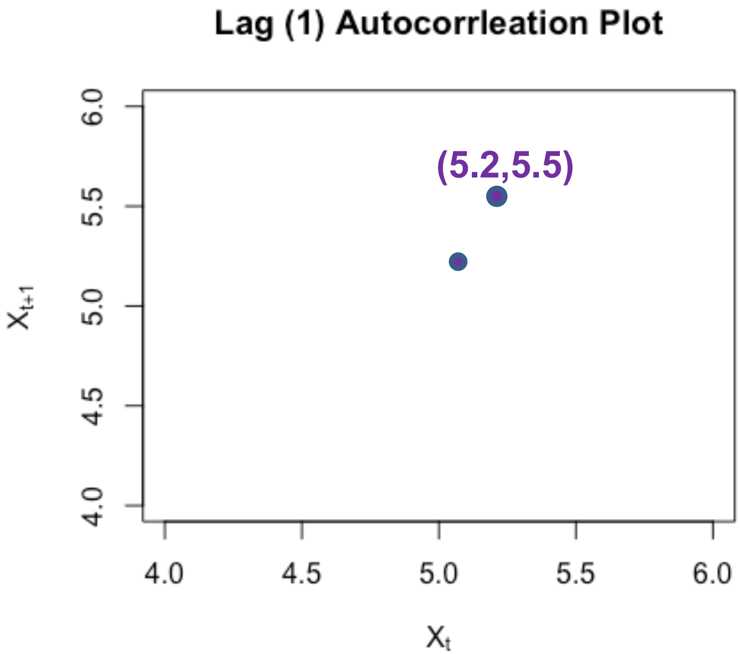
Example: Lag(1) Autocorrelation



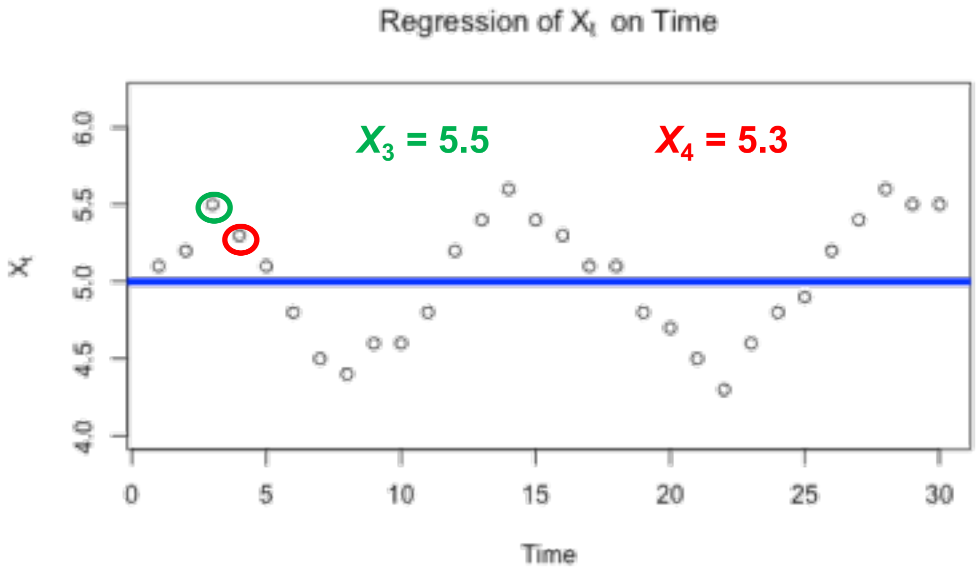


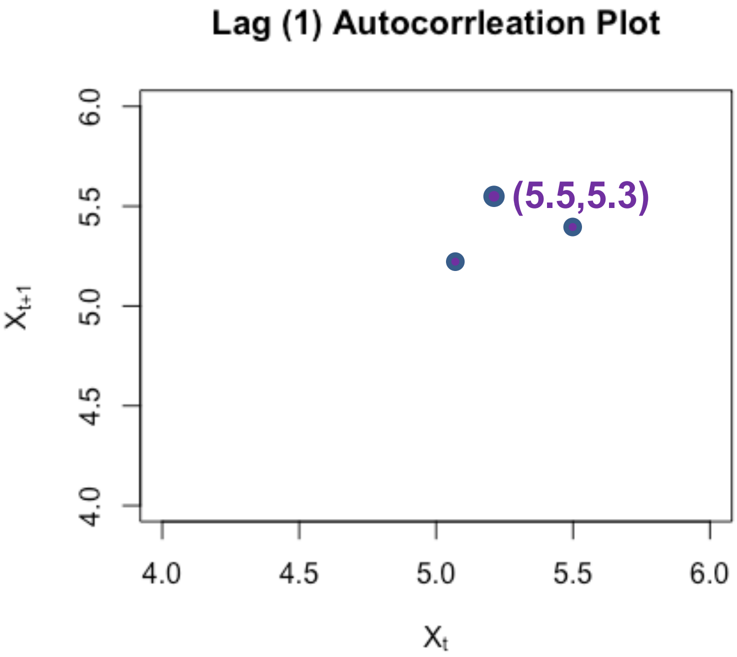
Example: Lag(1) Autocorrelation



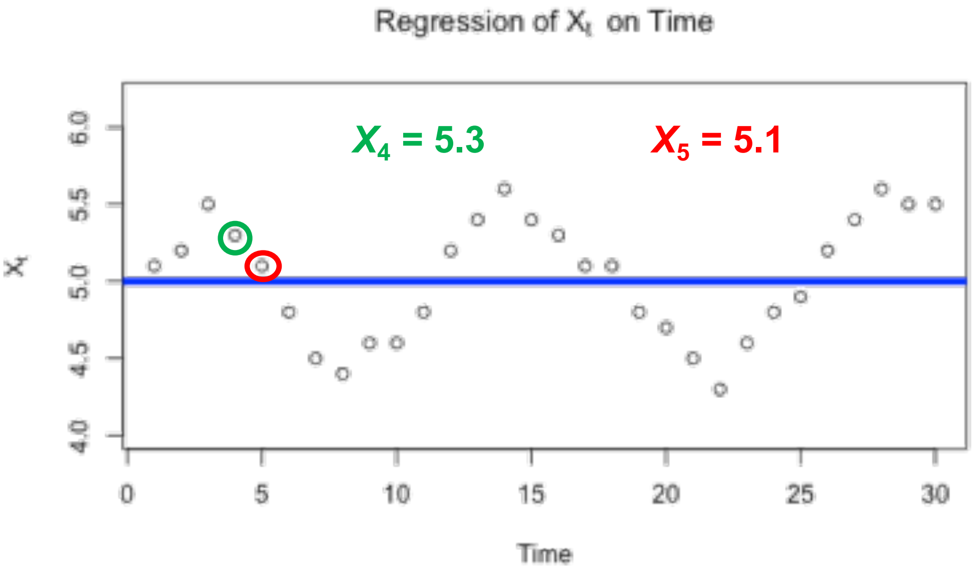


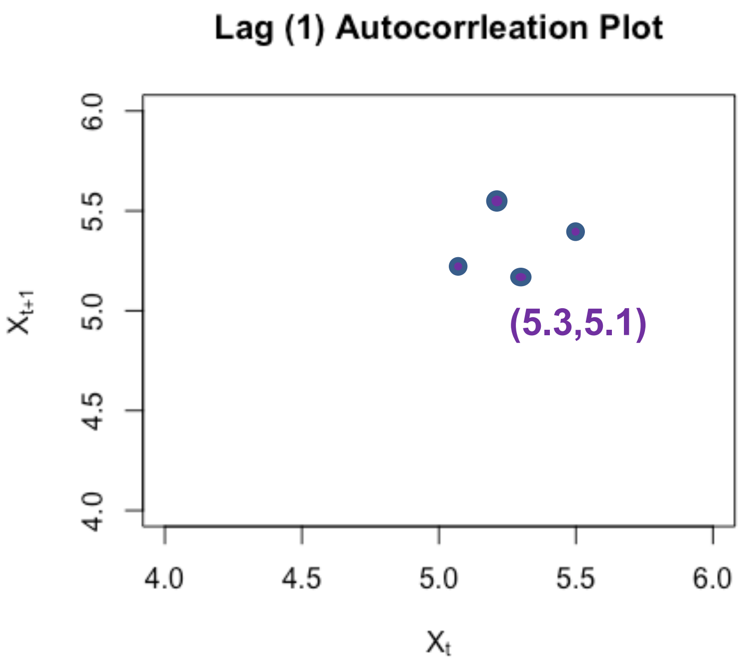
Example: Lag(1) Autocorrelation



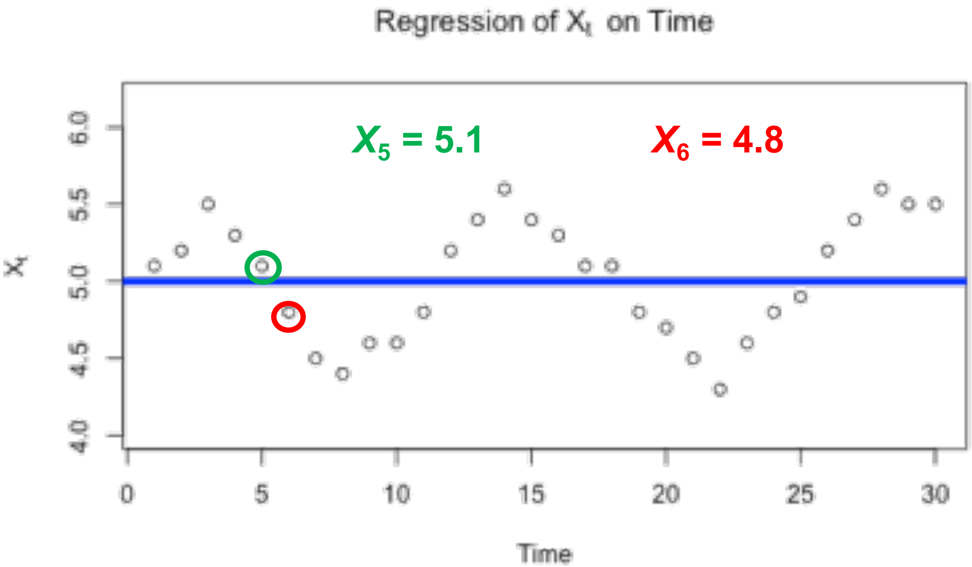


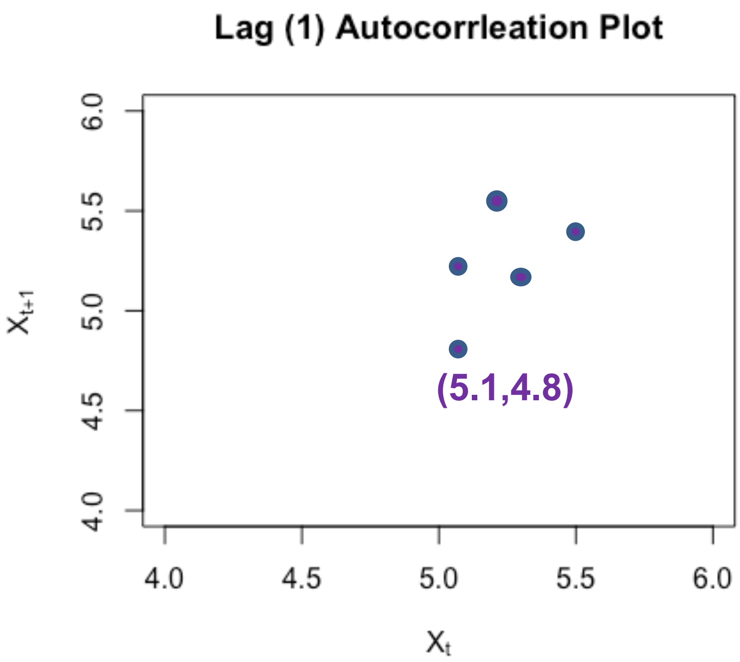
Example: Lag(1) Autocorrelation



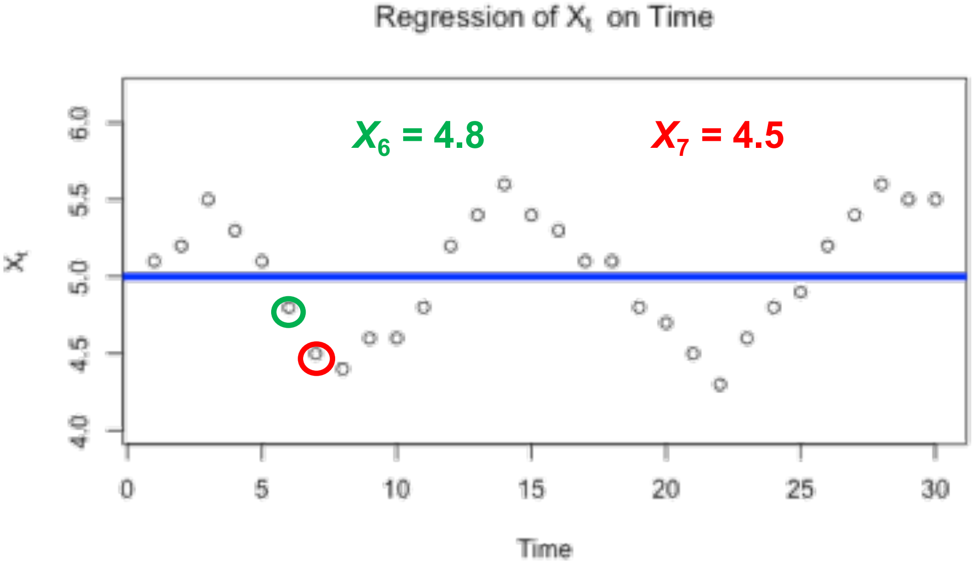


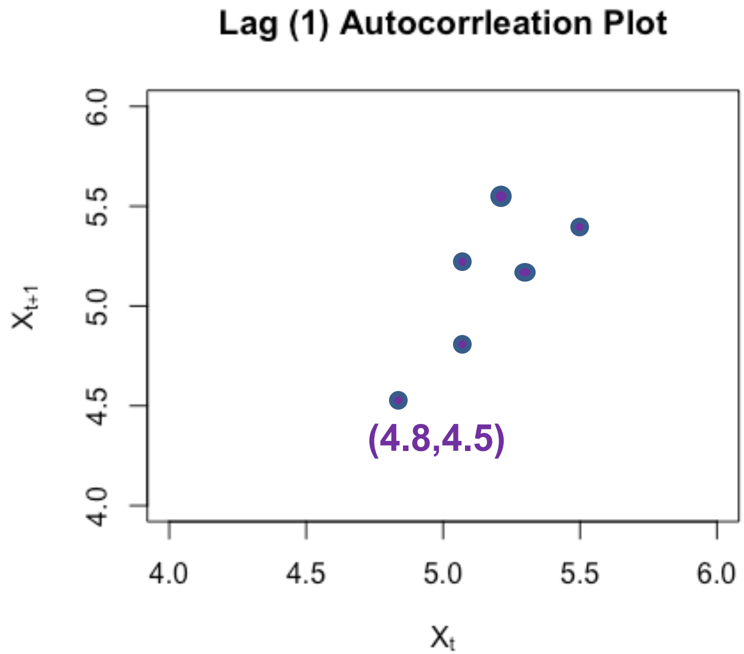
Example: Lag(1) Autocorrelation



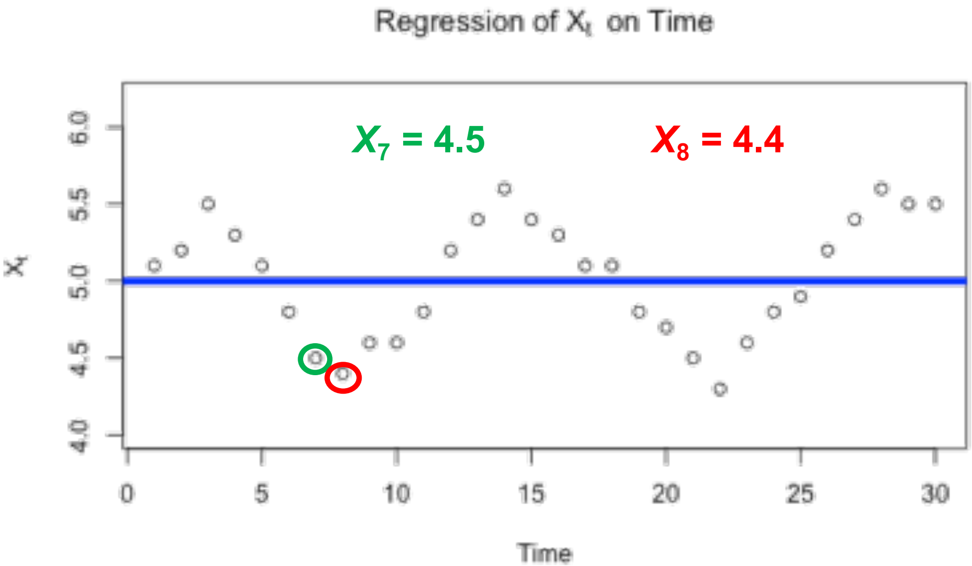


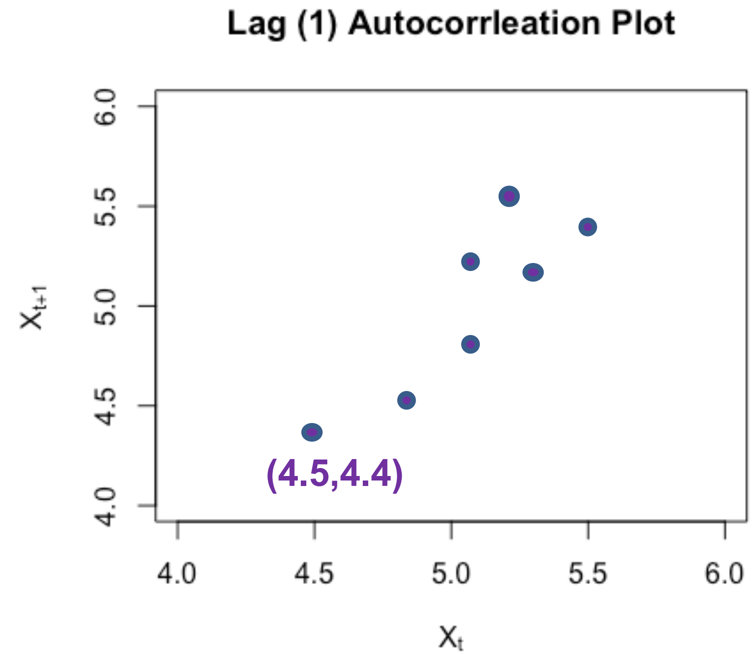
Example: Lag(1) Autocorrelation



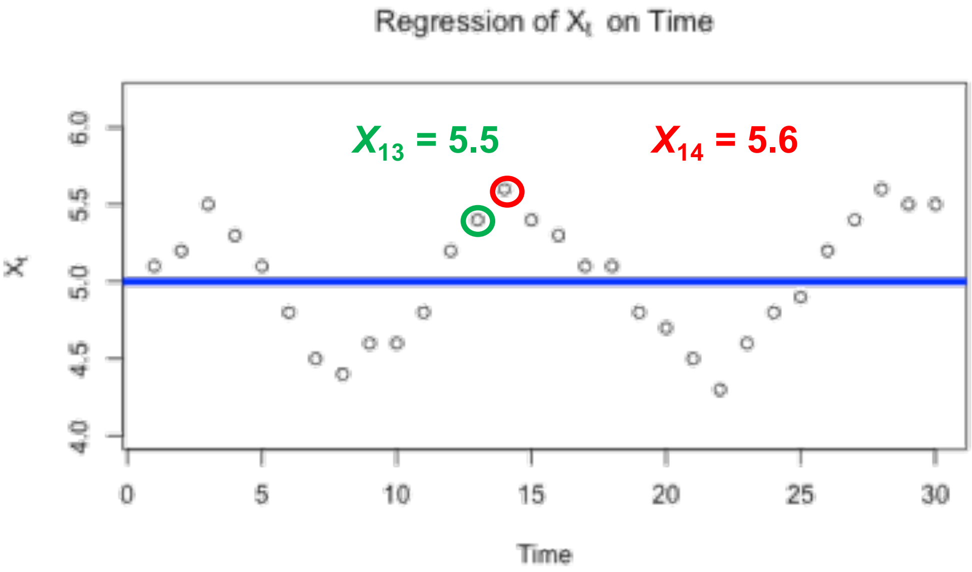


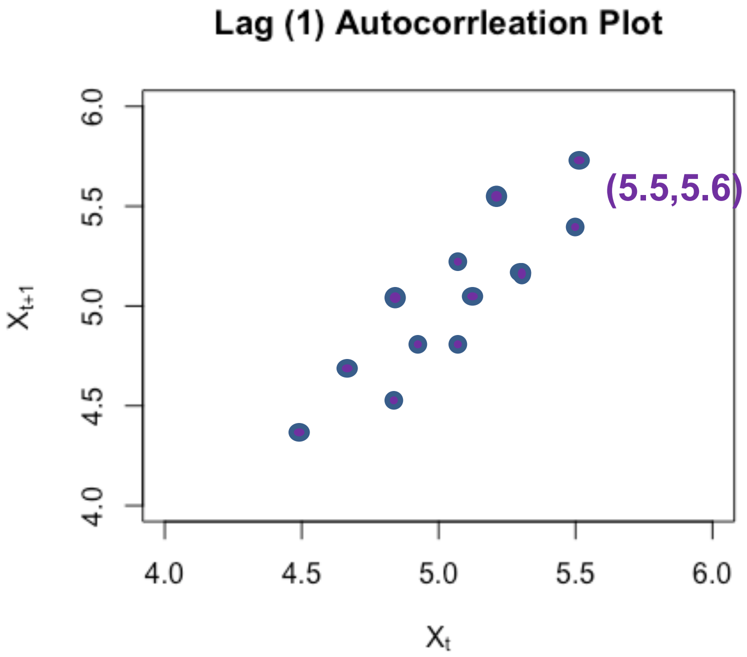
Example: Lag(1) Autocorrelation



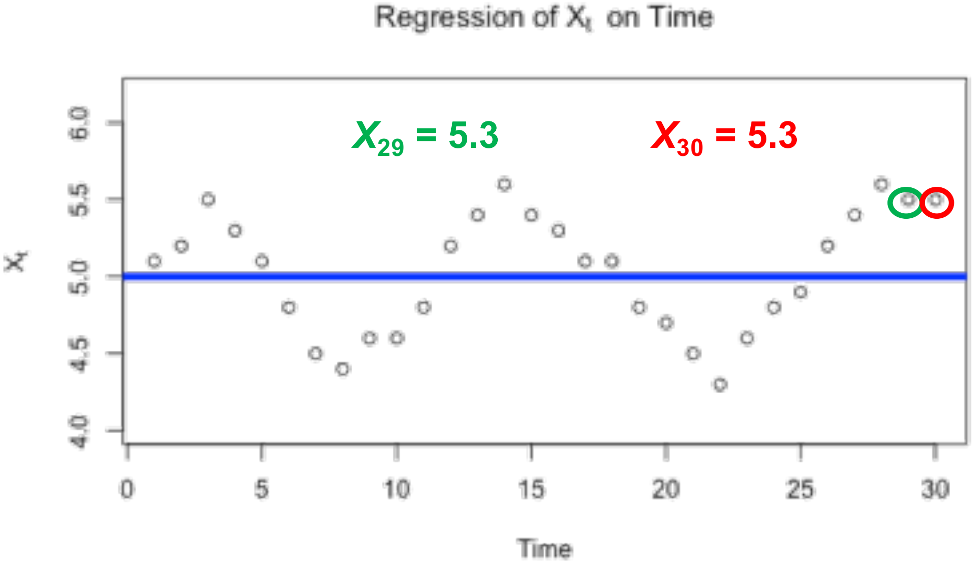


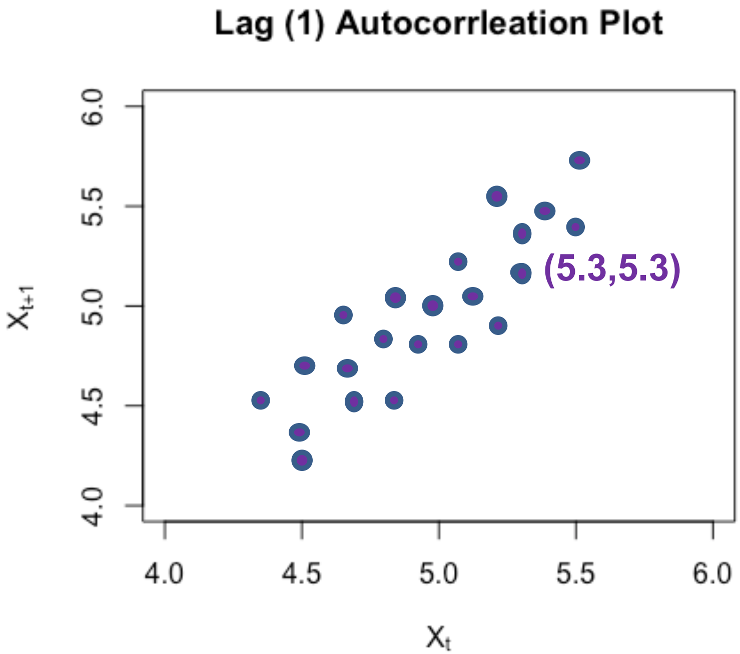
Example: Lag(1) Autocorrelation





Example: Lag(1) Autocorrelation



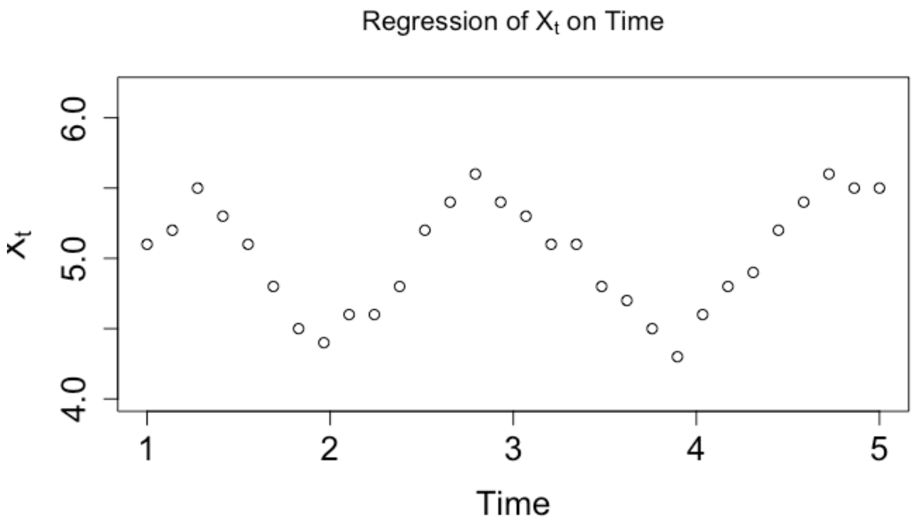


There is strong evidence that X_t and $X_t + 1$ are positively correlated and thus not independent.

Autocorrelation of Time Series with Dependent Observations (Lag = 1)

```
Y5 =
c(5.1,5.2,5.5,5.3,5.1,4.8,4.5,4.4,4.6,4.6,4.8,5.2,5.4,5.6,5.4,5.3,5.
1,5.1,4.8,4.7,4.5,4.3,4.6,4.8,4.9,5.2,5.4,5.6,5.5,5.5)
Time = seq(1,5,length = 30)
plot(Time,Y, main = "Regression of Y on Time",ylim = c(4,6.2))
```

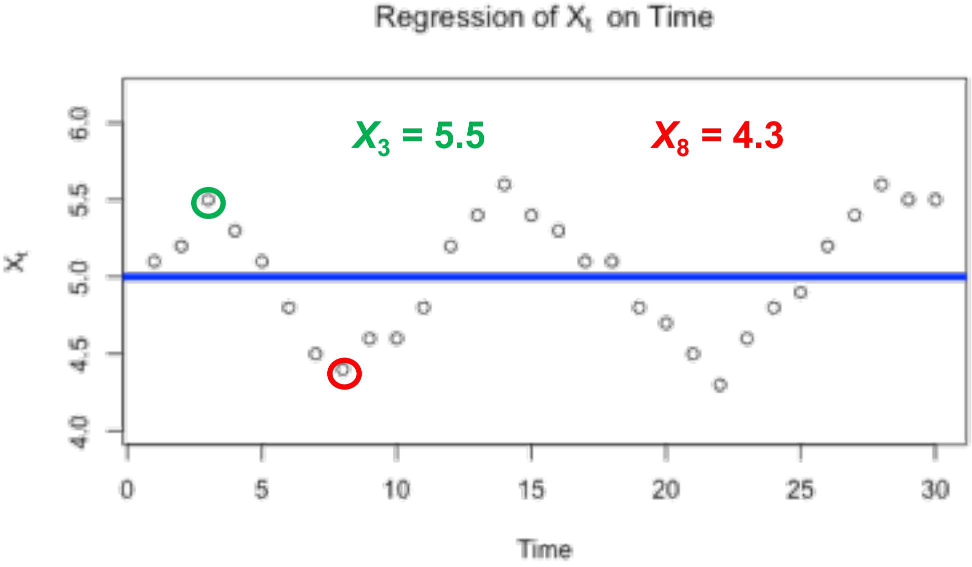
```
acf(Y5,plot = FALSE, lag.max = 1)
```

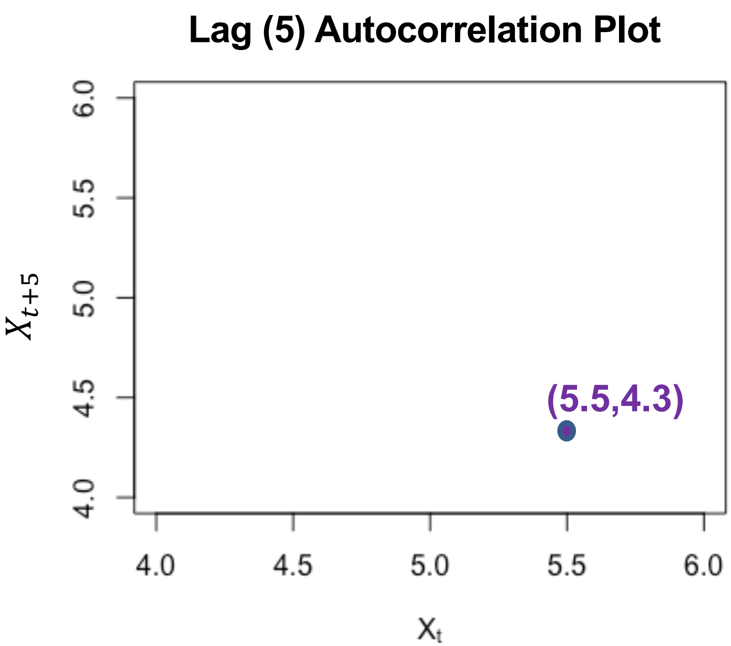



```
Autocorrelations of series 'Y5', by lag

    0     1
1.000 0.827
```

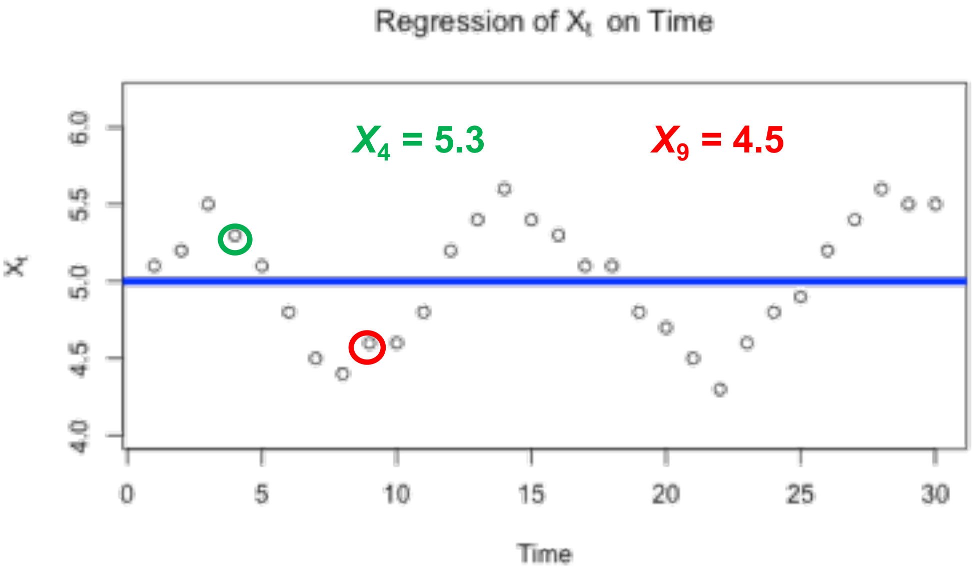
$$\hat{\rho}_{X_t,X_{t+1}} = \hat{\rho}_1 = r_1 = .827$$

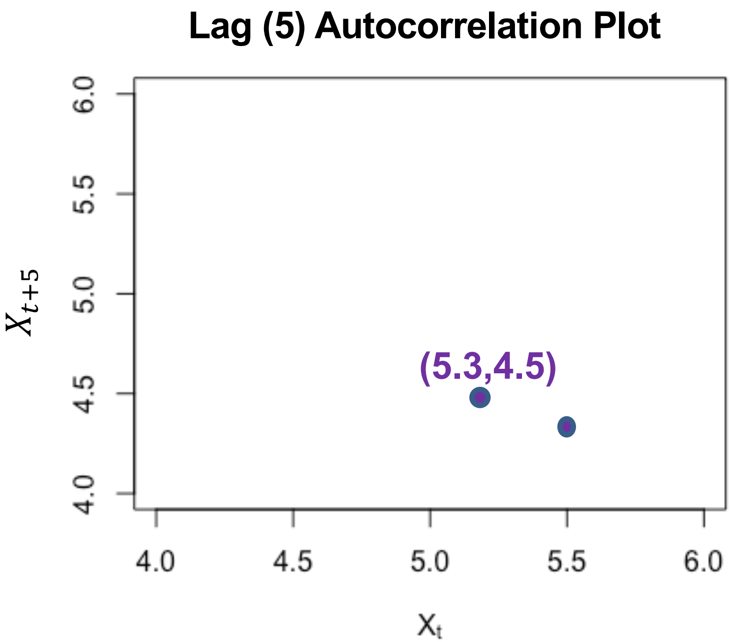
Example: Lag(5) Autocorrelation



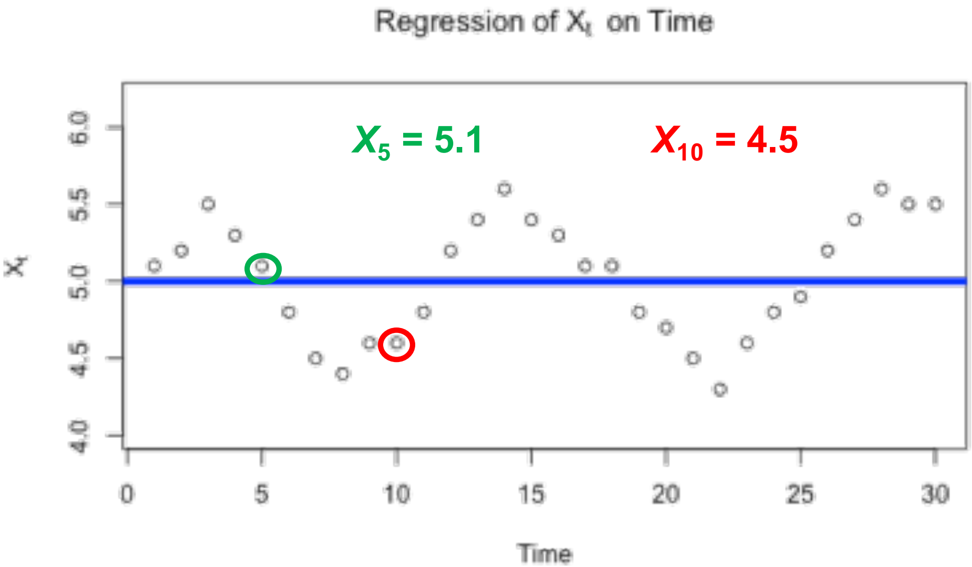


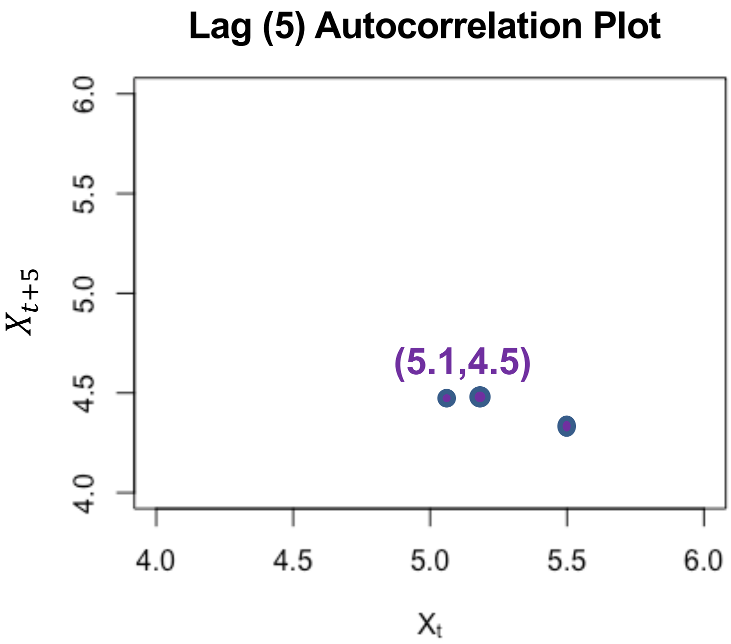
Example: Lag(5) Autocorrelation



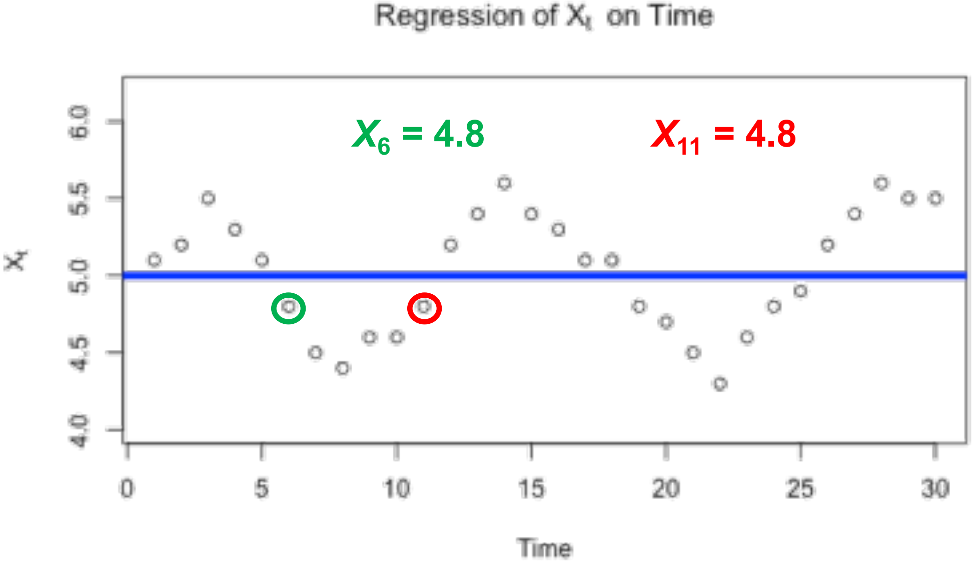


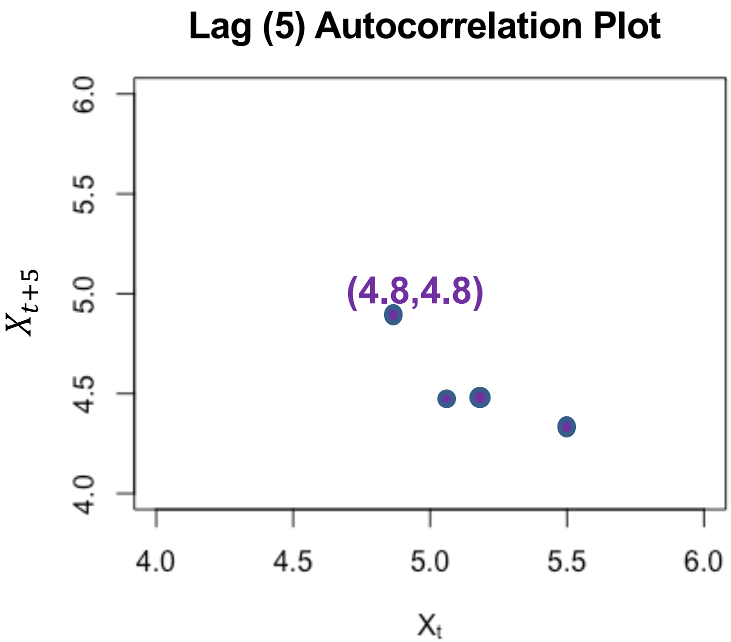
Example: Lag(5) Autocorrelation



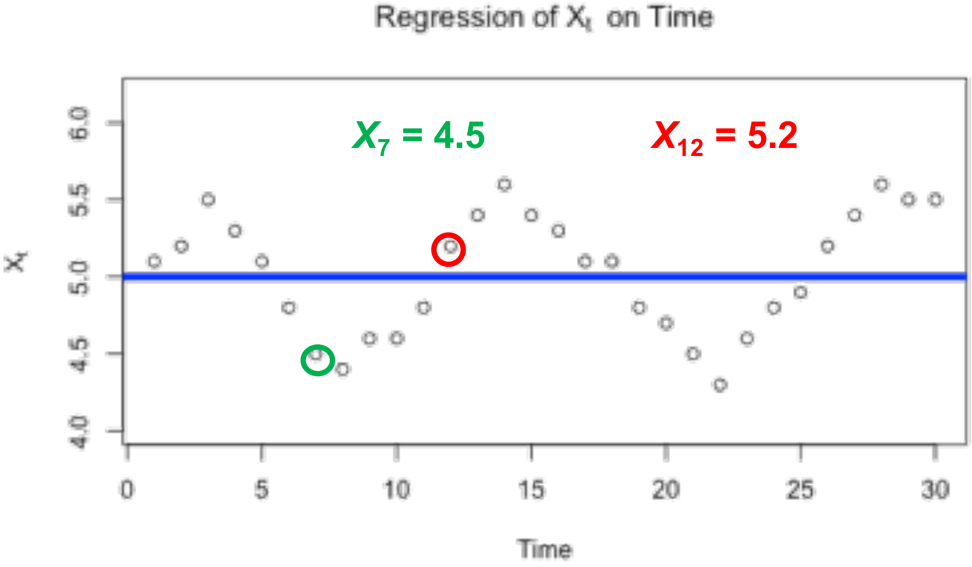


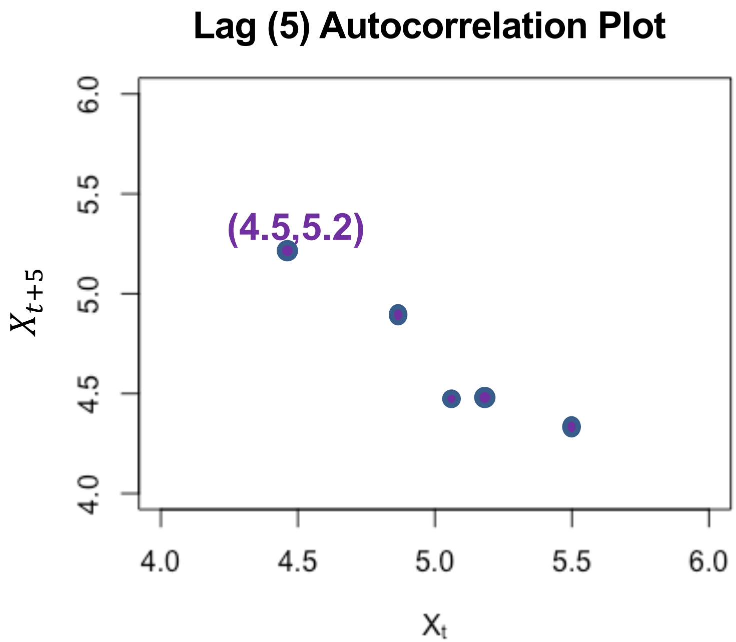
Example: Lag(5) Autocorrelation



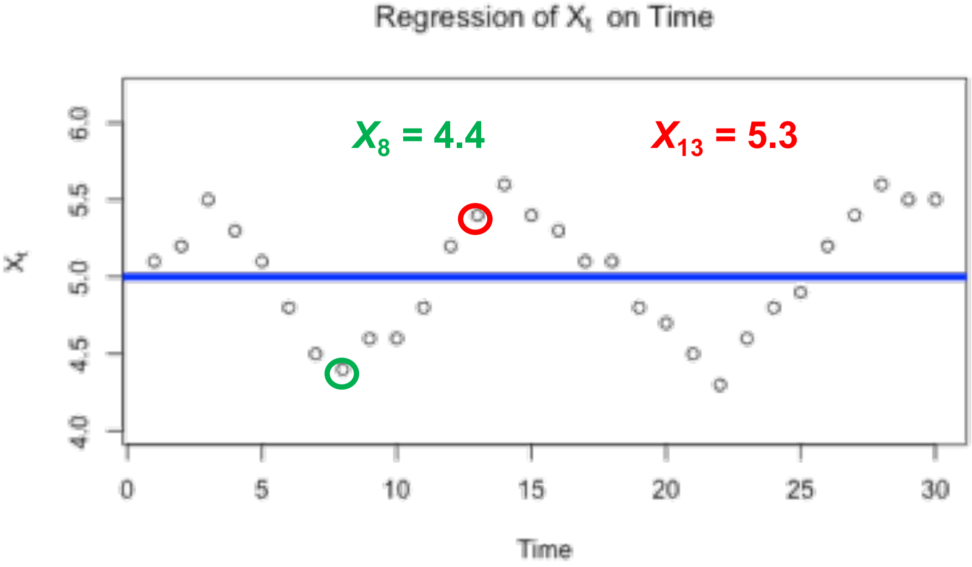


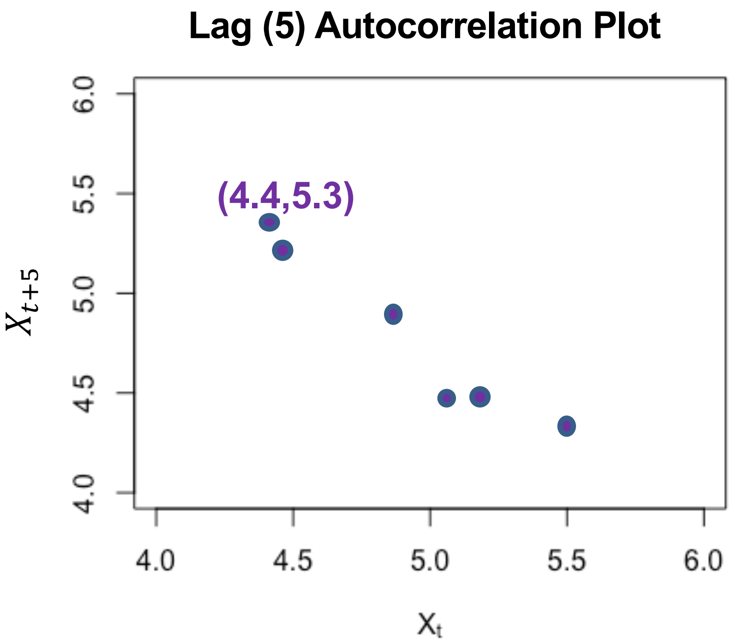
Example: Lag(5) Autocorrelation



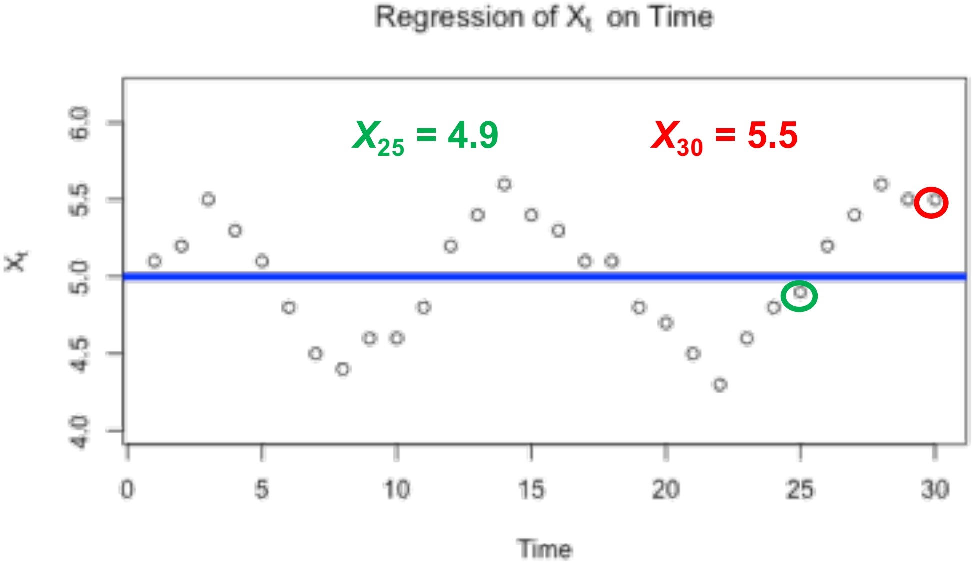


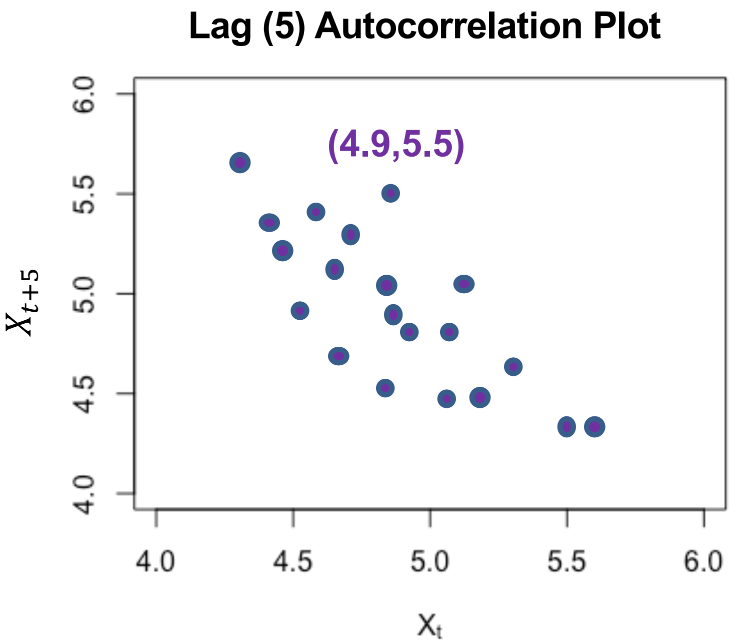
Example: Lag(5) Autocorrelation





Example: Lag(5) Autocorrelation

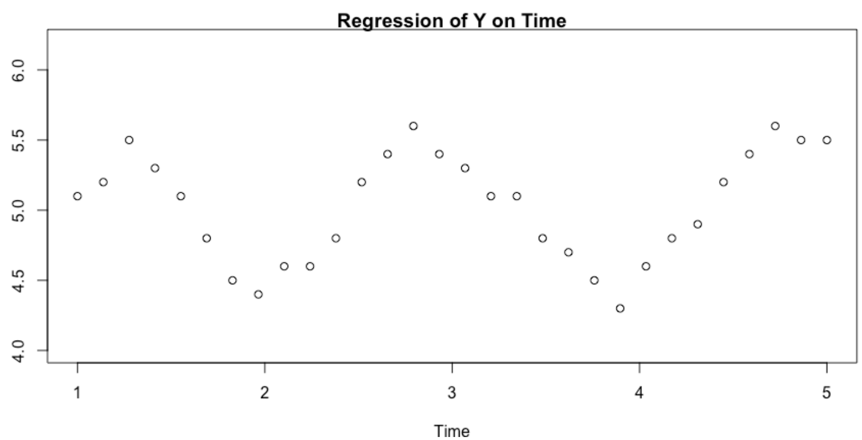




There is significant evidence that X_t and X_{t+5} are negatively correlated and thus also not independent.

Autocorrelation of Time Series with Dependent Observations with tswge (Lag = 5)

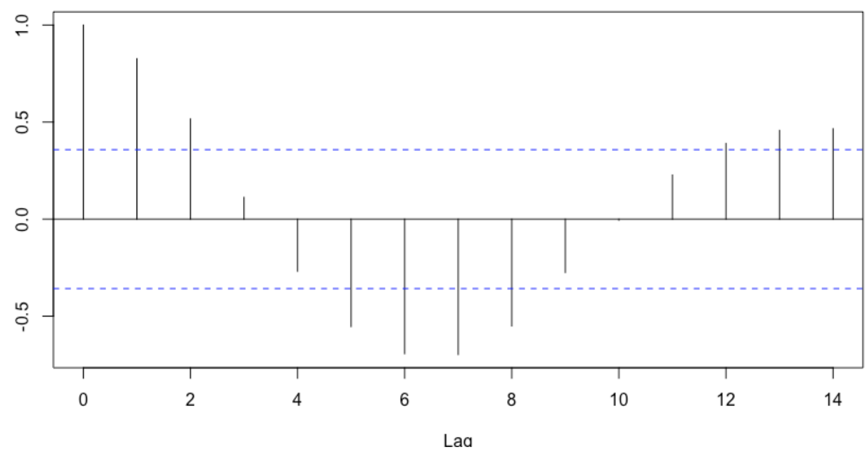
```
Y5 = c(5.1,5.2,5.5,5.3,5.1,4.8,4.5,4.4,4.6,4.6,4.8,5.2,5.4,5.6,5.4,5.3,5.1,5.1,4.8,4.7,4.5,4.3,4.6,4.8,4.9,5.2,5.4,5.6,5.5,5.5)

Time = seq(1,5,length = 30)

plot(Time,Y, main = "Regression of Y on Time",ylim = c(4,6.2), cex.axis = 1, )
```



acf(Y5,plot = FALSE, lag.max = 20)

```
Autocorrelations of series 'Y5', by lag

     0      1      2      3      4      5      6      7      8      9     10
 1.000  0.827  0.517  0.113 -0.269 -0.554 -0.694 -0.698 -0.551 -0.275 -0.006
    11     12     13     14     15     16     17     18     19     20
 0.227  0.390  0.457  0.466  0.365  0.174 -0.045 -0.223 -0.331 -0.330
```

acf(Y5)



Independence and Correlation

- **Theorem: If two events are independent, their correlation is 0.**

Recall that ρ is the population linear correlation coefficient.

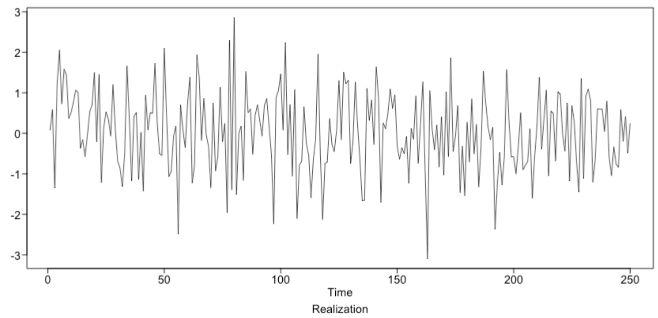
That is, if X_t and X_{t+k} are independent, $\rho_{X_t,X_{t+k}} = 0$.

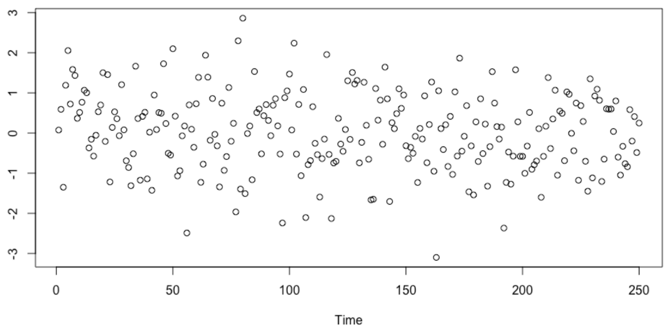
- **Corollary: If the correlation between two variables is nonzero, they are not independent.**

That is, if $\rho_{X_t,X_{t+k}} \neq 0$, then X_t and X_{t+k} are not independent.

Autocorrelation of Time Series with Independent Observations with tswge (Lag = 1)

```
> Realization1 = gen.arma.wge(n = 250)
>
> Time = seq(1,250,length.out = 250)
> plot(Time,Realization1)
```



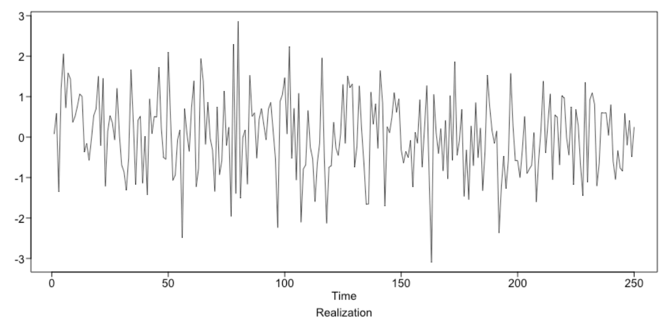


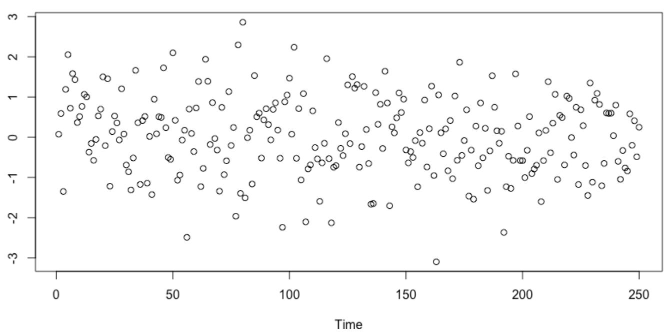

```
> acf(Realization1,plot = FALSE,lag.max = 1)

Autocorrelations of series 'Realization1', by lag

    0      1
1.000 -0.043
```

$$\hat{\rho}\,(t, t+1) = \hat{\rho}_1 = r_1 = -.043$$

Autocorrelation of Time Series with Independent Observations with tswge (Lag = 1 to *k*)

```
> Realization1 = gen.arma.wge(n = 250)
>
> Time = seq(1,250,length.out = 250)
> plot(Time,Realization1)
```

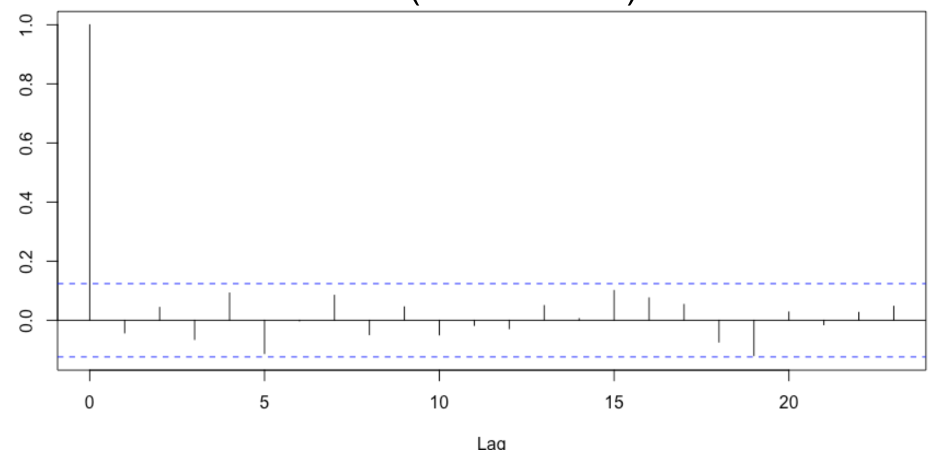




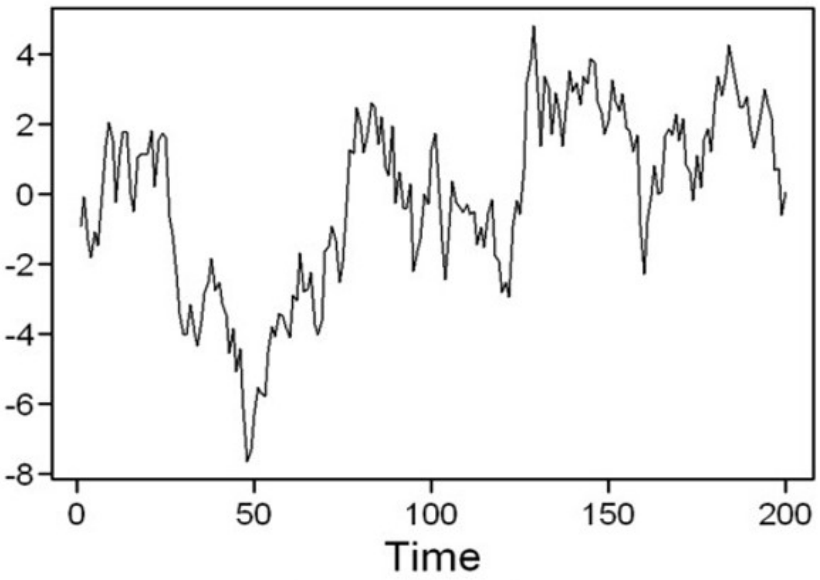
acf(Realization1,plot = FALSE,lag.max = 20)

```
Autocorrelations of series 'Realization1', by lag

     0      1      2      3      4      5      6      7      8      9     10
 1.000 -0.043  0.044 -0.065  0.092 -0.112 -0.002  0.084 -0.049  0.045 -0.050
    11     12     13     14     15     16     17     18     19     20
-0.018 -0.029  0.050  0.006  0.101  0.076  0.054 -0.074 -0.119  0.028
```

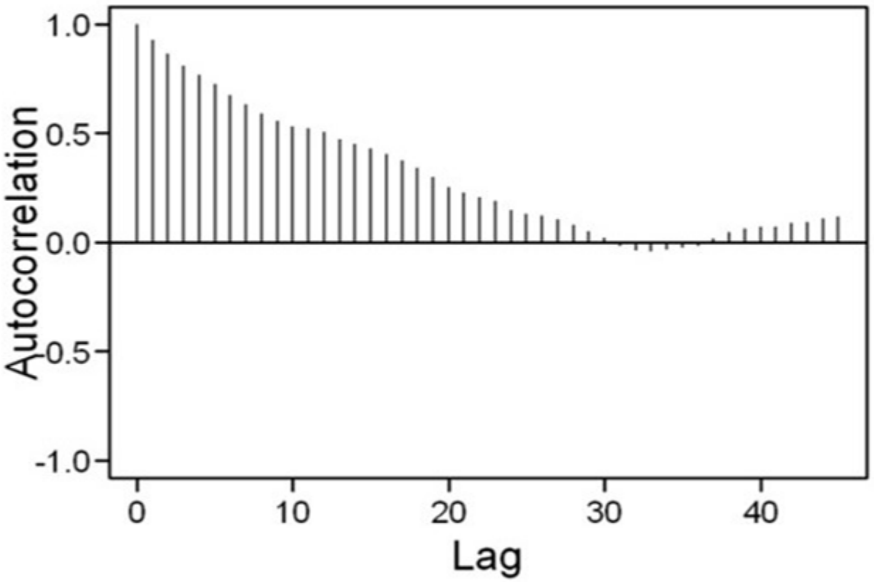
acf(Realization1)



Provide Example of Realizations and ACFS

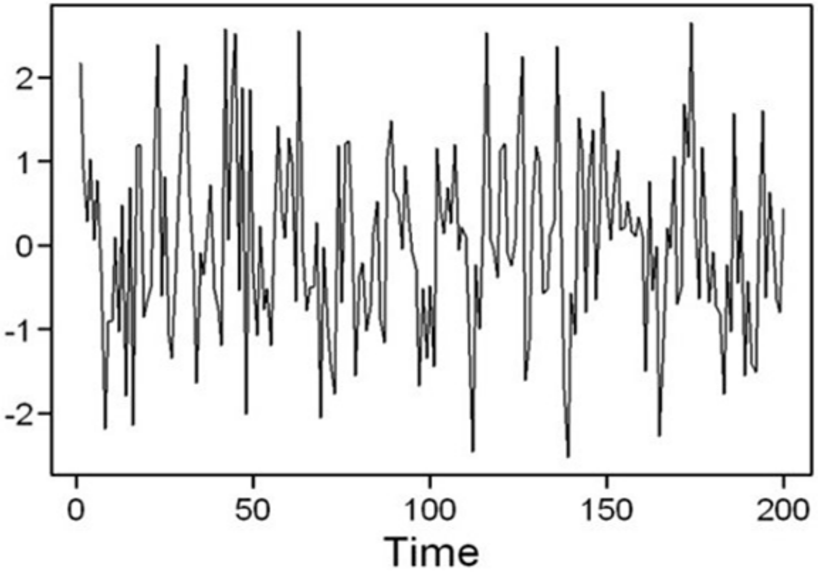


Realization

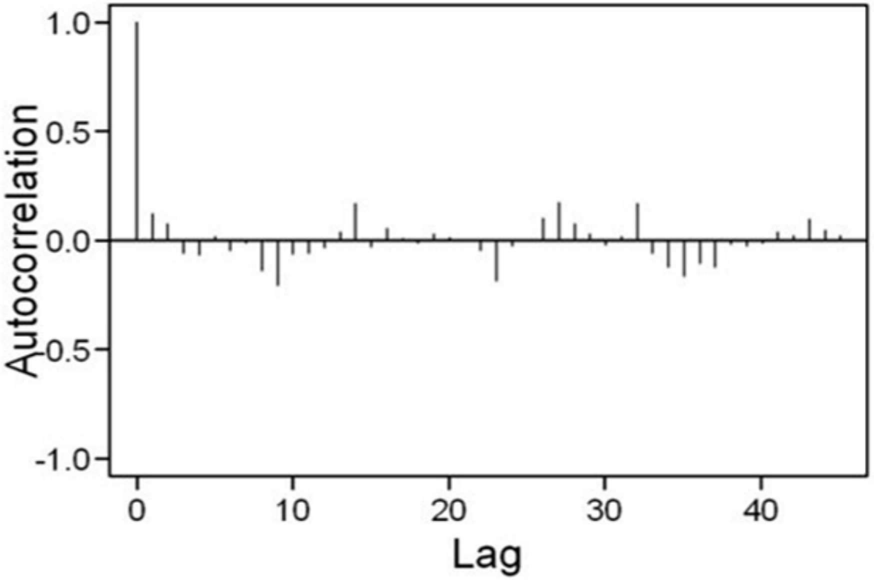


Sample autocorrelations*

*We will show how to estimate the autocorrelations in the next section.

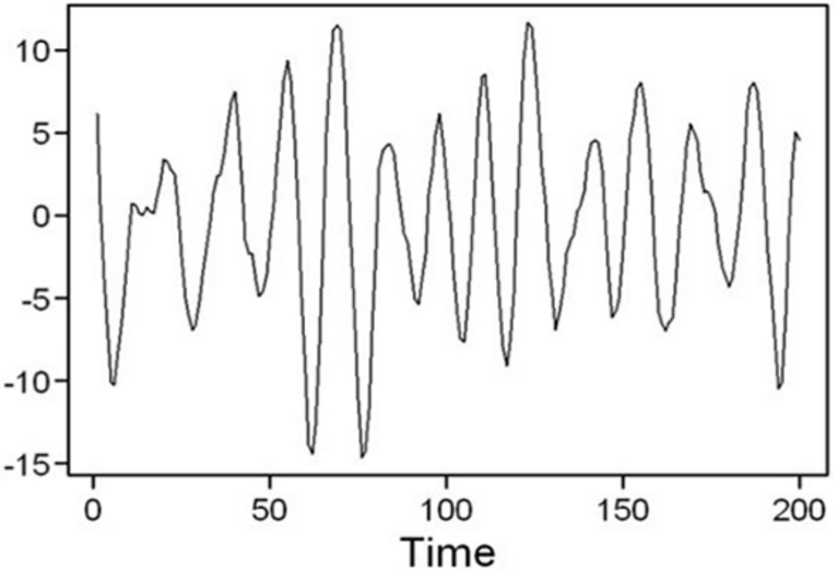


Realization

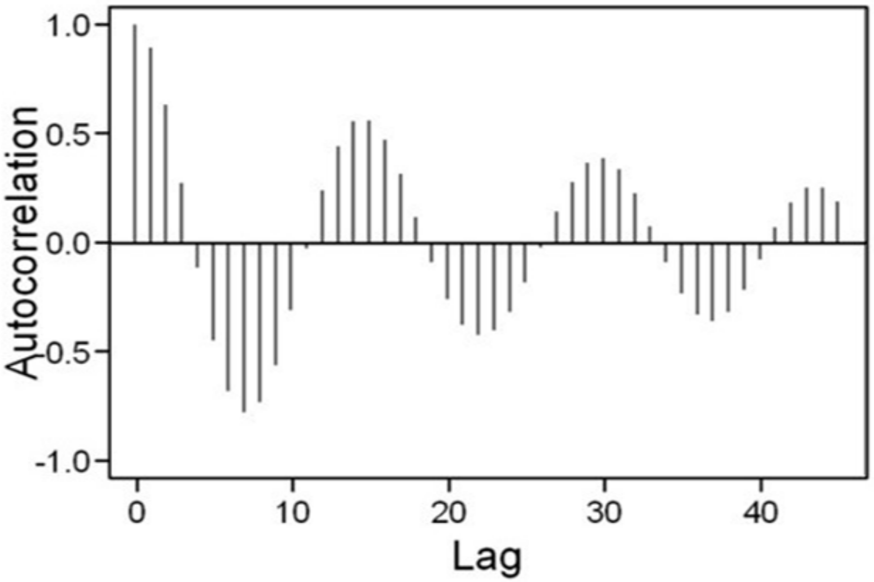


Sample autocorrelations*

*We will show how to estimate the autocorrelations in the next section.

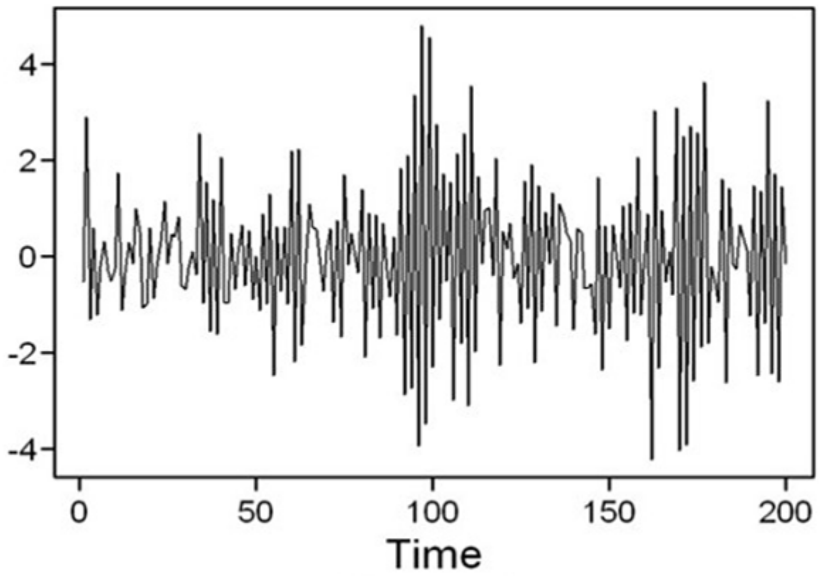


Realization

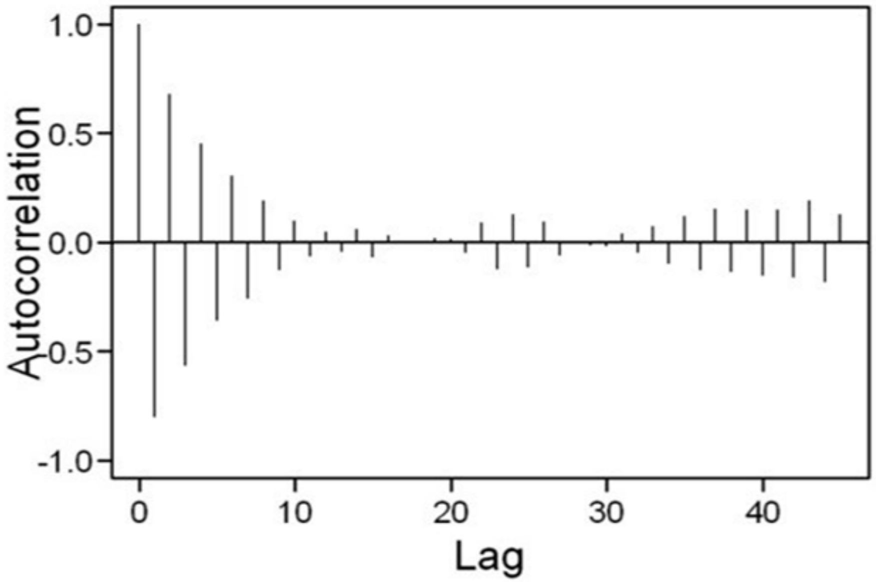


Sample autocorrelations*

*We will show how to estimate the autocorrelations in the next section.

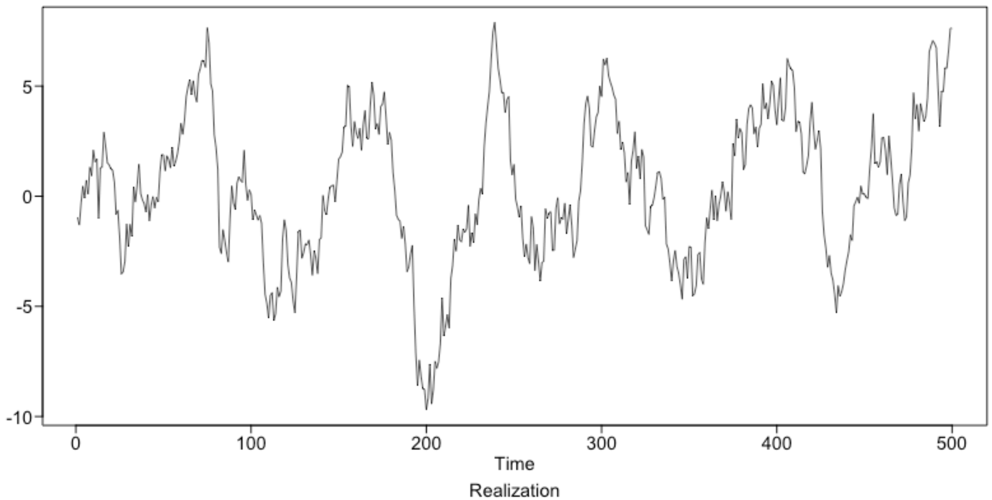


Realization



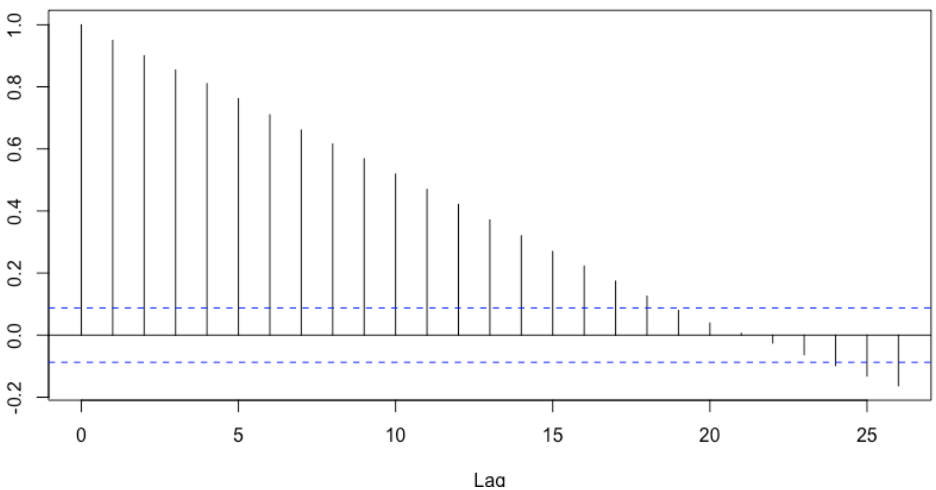
Sample autocorrelations*

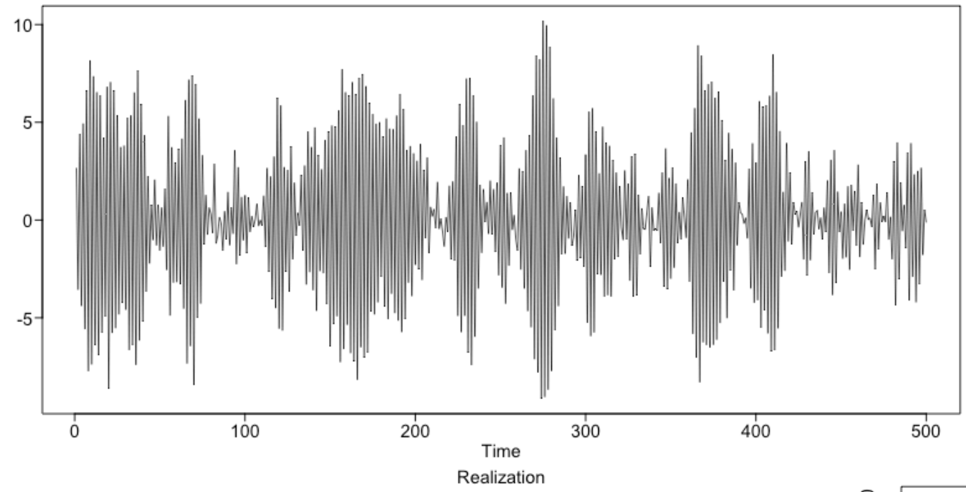
*We will show how to estimate the autocorrelations in the next section.



bb = gen.arma.wge(500,.96,sn = 5)

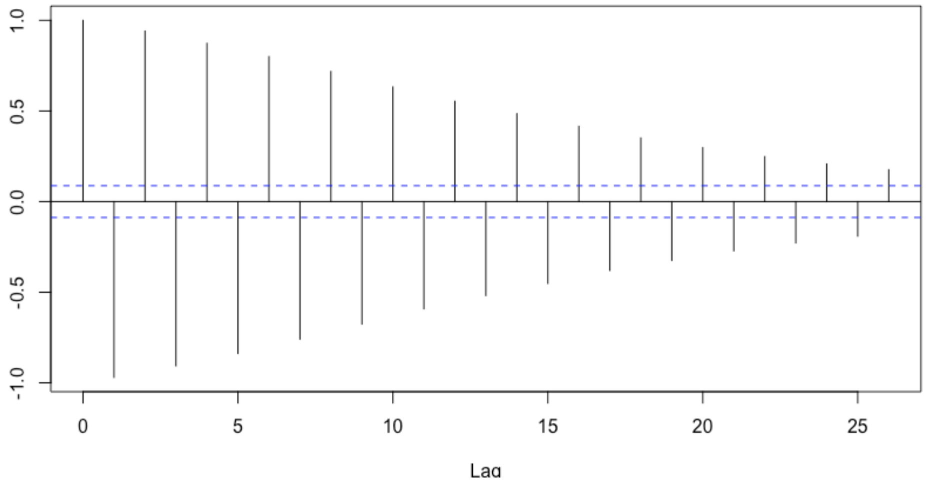
acf(bb)





aa = gen.arma.wge(500,–.96,sn = 2)

acf(aa)



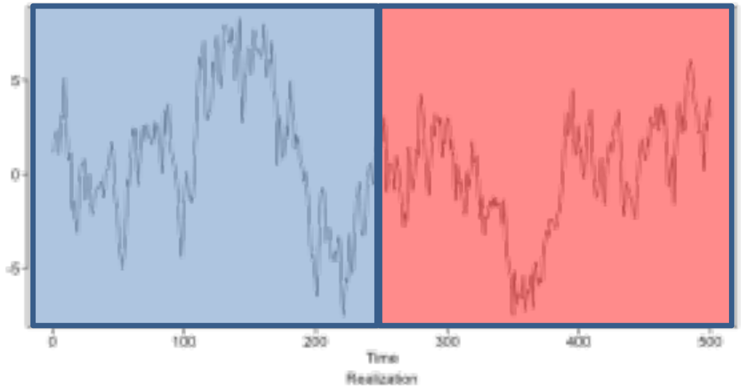
DataScience@SMU

Stationary Covariance

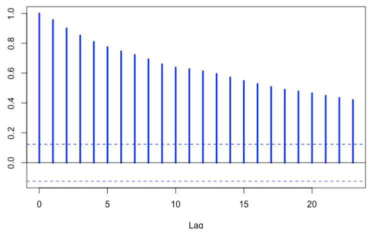
Weak Stationary Process: Requirement 3

3. The correlation of X_{t_1} and X_{t_2} depends only on $t_2 - t_1$. That is, the correlation between data points is dependent only on how far apart they are, not where they are.

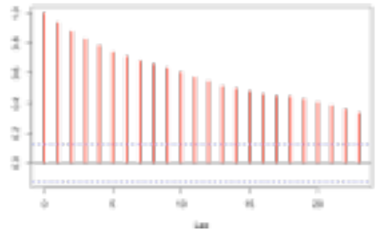
$$Notation: Cor(X_t, X_{t+h}) = \rho_h$$



Realization = gen.arma.wge(500,.95,0,plot = TRUE,sn = 784)



acf(Realization[1:250],plot = TRUE)

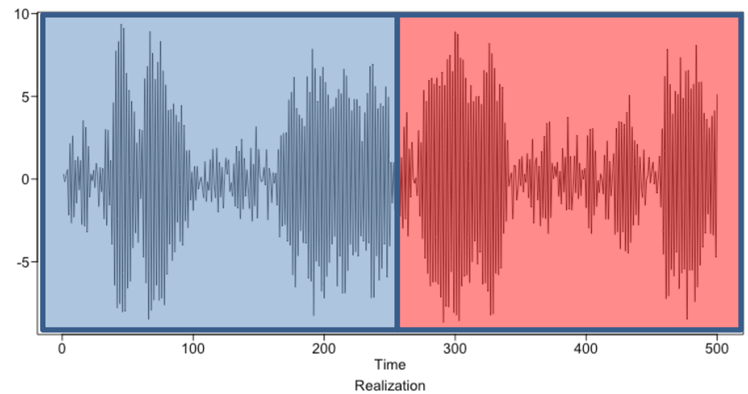


acf(Realization[251:500],plot = TRUE)

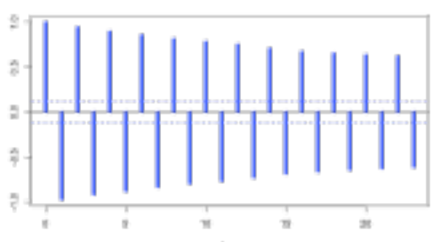
Weak Stationary Process: Requirement 3

3. The covariance of X_{t_1} and X_{t_2} depends only on $t_2 - t_1$. That is, the covariance between data points is dependent only on how far apart they are, not where they are.

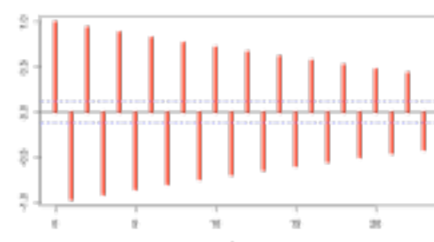
$$Notation: Cor(X_t, X_{t+h}) = \rho_h$$



Realization = gen.arma.wge(500,-.95,0,plot = TRUE,sn = 45)



acf(Realization[1:250],plot = TRUE)

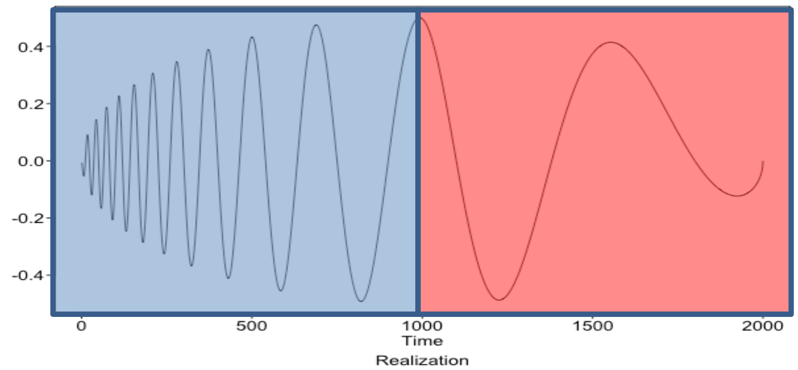


acf(Realization[251:500],plot = TRUE)

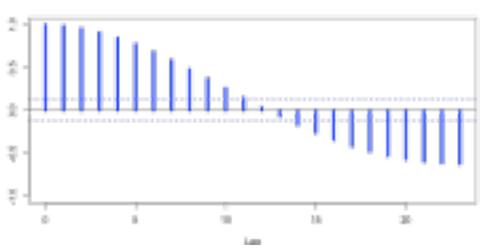
Weak Stationary Process: Requirement 3

3. The covariance of X_{t_1} and X_{t_2} depends only on $t_2 - t_1$. That is, the covariance between data points is dependent only on how far apart they are, not where they are.

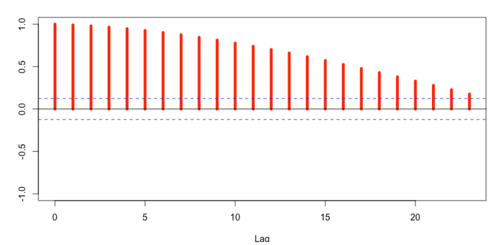
$$Notation: Cor(X_t, X_{t+h}) = \rho_h$$



doppler data



acf(doppler[1:251],plot = TRUE,ylim = c(–1,1))



acf(doppler[251:500],plot = TRUE,ylim = c(–1,1))

DataScience@SMU

Stationary Time Series

Stationary Process

1. Subpopulations of *X* for a given time have constant mean for all *t*.

$$E[X_t] = \mu$$

2. Subpopulations of *X* for a given time have a constant variance for all *t*.

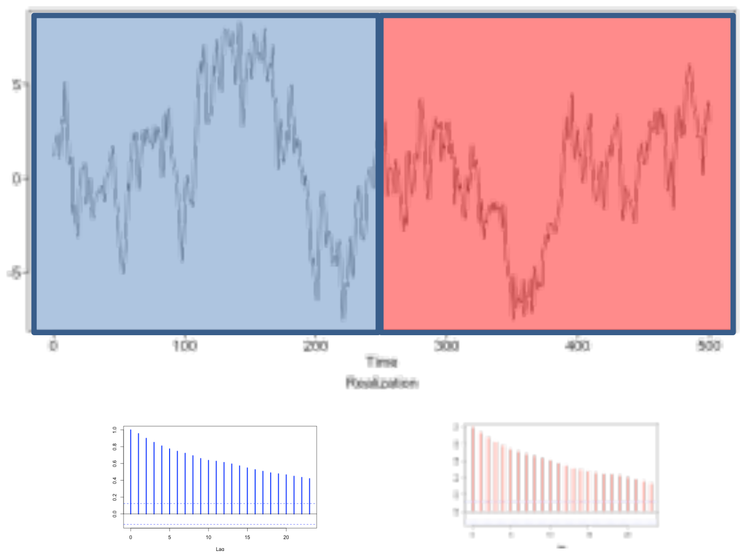
$$\text{Var}[X_t] = \sigma^2$$

3. The correlation of X_{t_1} and X_{t_2} depends only on $t_2 - t_1$. That is, the covariance between data points is dependent only on how far apart they are, not where they are.

$$Notation: Cor(X_t, X_{t+h}) = \rho_h$$

*See the text for the definition of strict stationarity. The use of the term “stationary” in this course will refer to weak stationarity.

Stationary Process



1. Subpopulations of *X* for a given time have constant mean for all *t*.

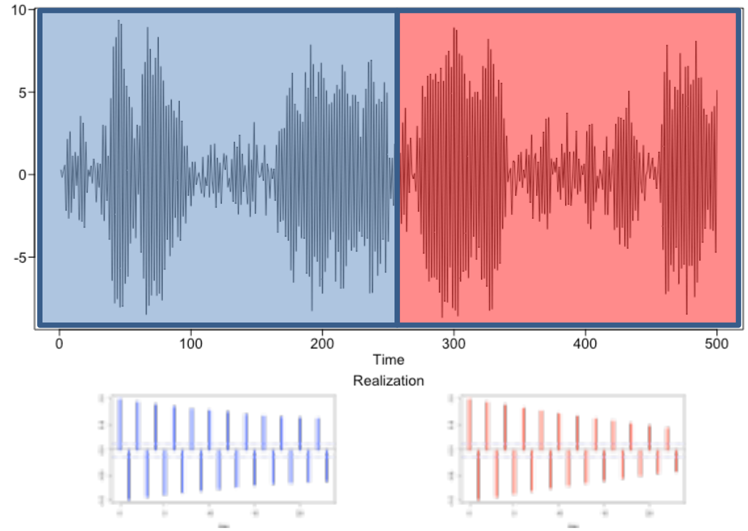
$$E[X_t] = \mu$$

2. Subpopulations of *X* for a given time have a constant and finite variance for all *t*.

$$\text{Var}[X_t] = \sigma^2 < \infty$$

3. The correlation of X_{t_1} and X_{t_2} depends only on $t_2 - t_1$. That is, the covariance between data points is dependent only on how far apart they are, not where they are.

Stationary Process



1. Subpopulations of *X* for a given time have constant mean for all *t*.

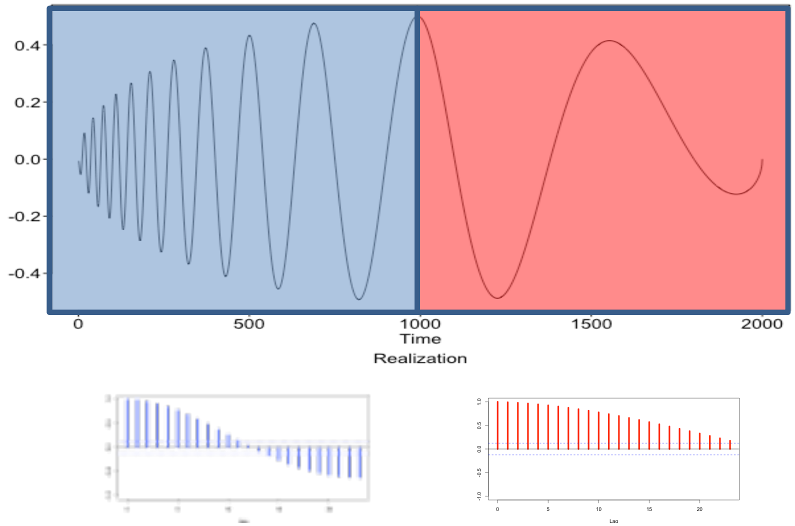
$$E[X_t] = \mu$$

2. Subpopulations of *X* for a given time have a constant and finite variance for all *t*. ?

$$\mathrm{Var}[X_t] = \sigma^2 < \infty$$

3. The correlation of X_{t_1} and X_{t_2} depends only on $t_2 - t_1$. That is, the covariance between data points is dependent only on how far apart they are, not where they are.

Stationary Process



1. Subpopulations of *X* for a given time have constant mean for all *t*.

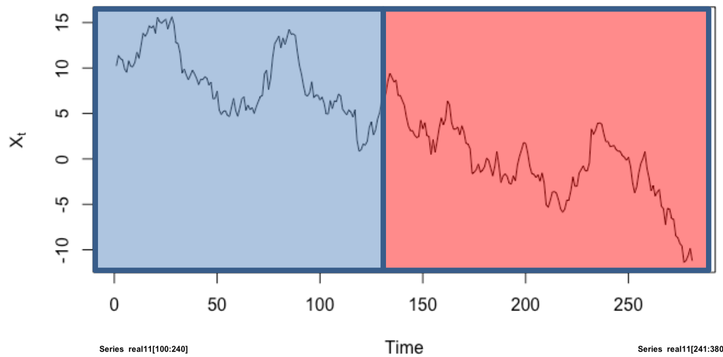
$$E[X_t] = \mu$$

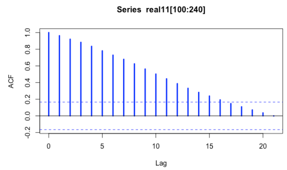
2. Subpopulations of *X* for a given time have a constant and finite variance for all *t*.

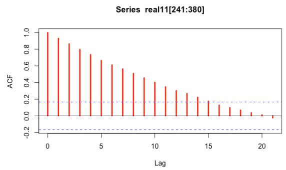
$$\text{Var}[X_t] = \sigma^2 < \infty$$

3. The correlation of X_{t_1} and X_{t_2} depends only on $t_2 - t_1$. That is, the covariance between data points is dependent only on how far apart they are, not where they are.

Stationary Process







1. Subpopulations of *X* for a given time have constant mean for all *t*. ?

$$E[X_t] = \mu$$

2. Subpopulations of *X* for a given time have a constant and finite variance for all *t*.

$$\text{Var}[X_t] = \sigma^2 < \infty$$

3. The correlation of X_{t_1} and X_{t_2} depends only on $t_2 - t_1$. That is, the covariance between data points is dependent only on how far apart they are, not where they are.

DataScience@SMU

Estimation

Introduction

Key Point About Stationary Processes for Which $\rho_k \rightarrow 0$

For stationary time series, if the autocorrelations approach 0 as the distance between time points increases, then:

A single realization can be used to estimate:

- The mean
- The variance
- Autocorrelation

DataScience@SMU

Estimating the Mean of a Stationary Series

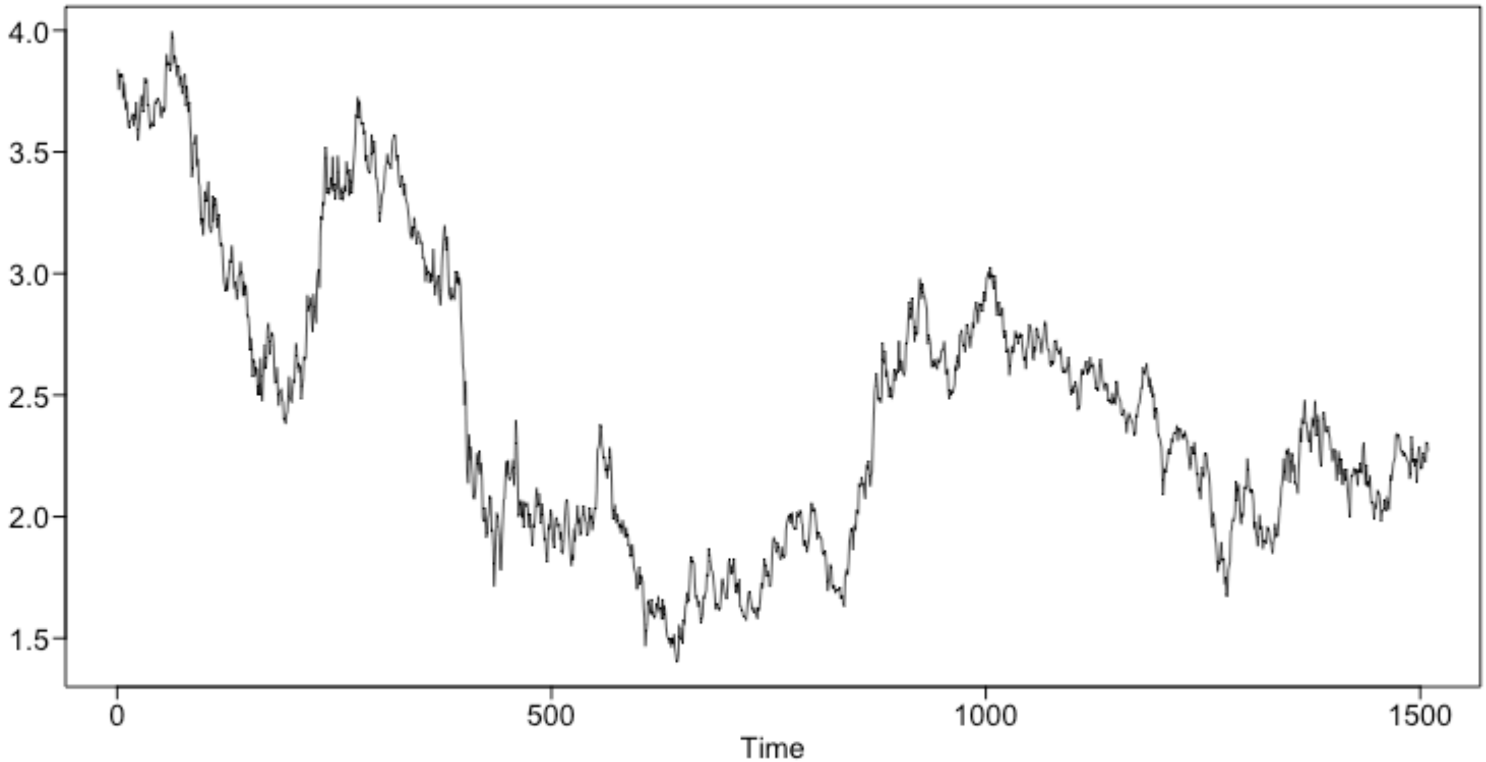
Typical Situation

- We only have one realization of finite length, n (i.e., $X_1, X_2, \ldots, X_n$).

If X_t is a discrete stationary time series, an unbiased estimate of μ is:

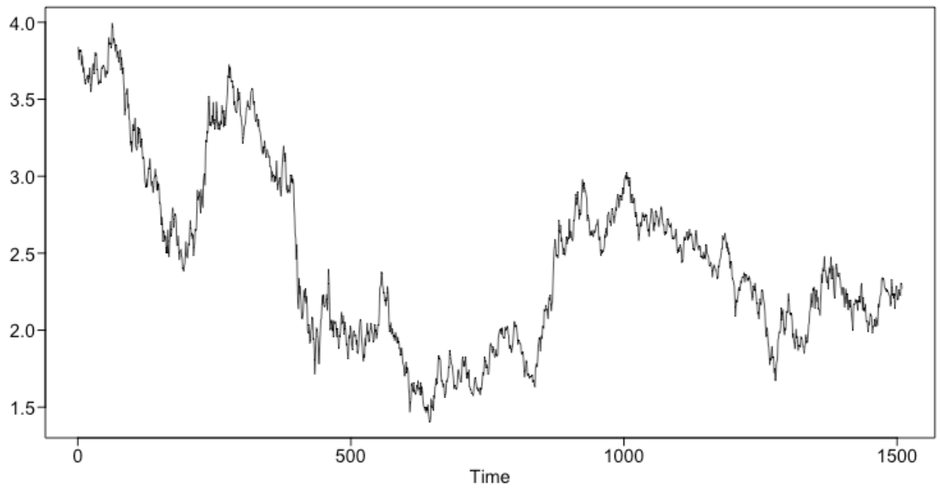
$$\bar{X}_t = \frac{\sum_{i=1}^{n} x_i}{n} \; for\ all\ t$$

Example: 10-Year U.S. Treasury Bond Rate



We will assume that this data is stationary.

Example: 10-Year U.S. Treasury Bond Rate



Assuming this data comes from a stationary process, we can estimate the mean bond rate with the sample mean of all the bond rates in the sample.

$$\bar{X}_t = \frac{\sum_{i=1}^{n} x_i}{n} = 2.457\%$$

assume bond data is in df called bond
mean(bond$Adj.Close

```
> mean(bond$Adj.Close)
[1] 2.457420808
```

DataScience@SMU

Estimating the Variance of a Stationary Series

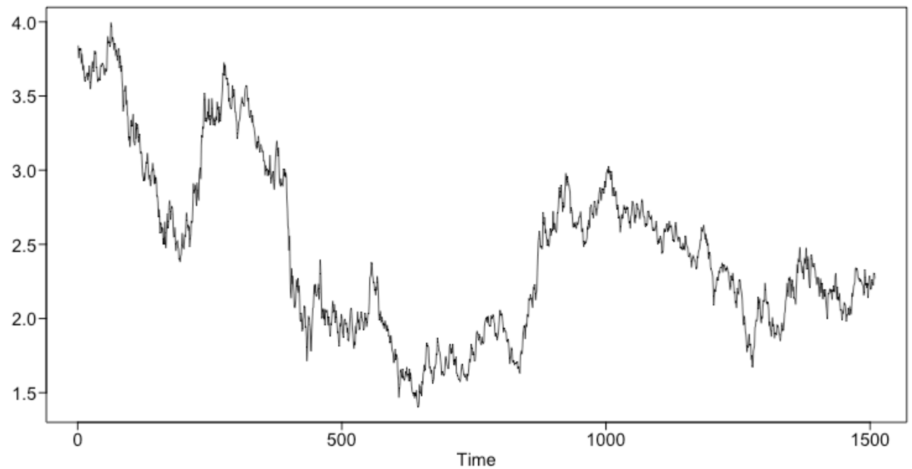
Theorem

If X_t is a discrete stationary time series, then the variance of $\bar{X}$ based on a realization of length *n* is given by:

$$Var(\bar{X}) = \frac{\sigma^2}{n} \sum_{k=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right) \rho_k$$

σ^2 will be estimated by the normal estimate of the variance using all the data, and we will provide further details on the estimation of ρ_k in the next section.

Example: 10-Year U.S. Treasury Bond Rate



Assuming this data comes from a stationary process, we can estimate the variance of the sample mean:

$$\widehat{Var(\bar{X})} = \frac{\hat{\sigma}^2}{n} \sum_{k=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right) \hat{\rho}_k$$

$$= \frac{\hat{\sigma}^2}{n} \left(1 + 2 * \sum_{k=1}^{n-1} \left(1 - \frac{|k|}{n}\right) \hat{\rho}_k \right)$$

=0.00358

```
xdf = read.csv(file.choose(),header = TRUE)
x = as.numeric(paste(xdf$Adj.Close))
x = x[!is.na(x)]
n=length(x) #n = 1509
nlag=1508 #n-1
m=mean(x)
v=var(x,na.rm = TRUE)
gamma0=var(x)*(n-1)/n
aut=acf(x,lag.max=1508) #n-1
sum=0
for  (k in 1:nlag) {sum=sum+(1-k/n)*aut$acf[k+1]*gamma0}
vxbar=2*sum/n+gamma0/n   #note the mult of sum by 2
vxbar
```

```
> vxbar
[1] 0.003580415821
```

DataScience@SMU

Confidence Intervals

Confidence Interval

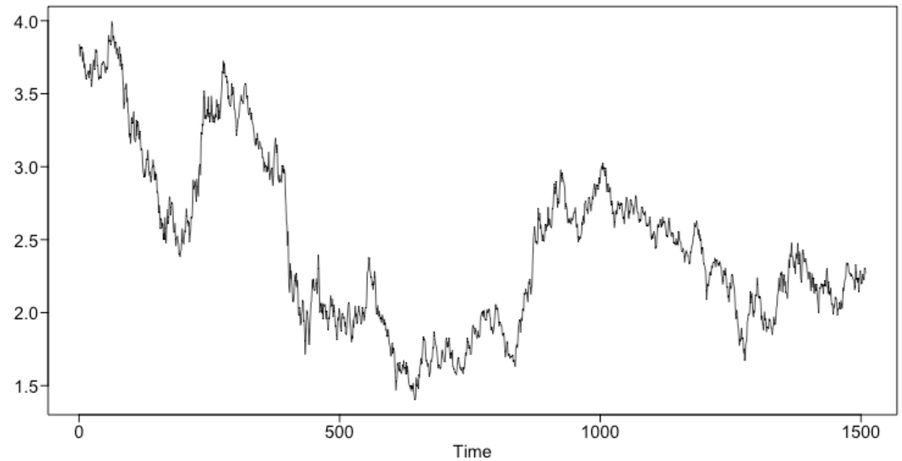
Once an estimate of the mean and variance we can obtain approximate 95% confidence intervals for from the familiar:

$$\bar{X} \pm 1.96 * \sqrt{\frac{\hat{\sigma}^2}{n} \sum_{k=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right) \hat{\rho}_k}$$

$$\bar{X} \pm 1.96 * \sqrt{\frac{\hat{\sigma}^2}{n} \left(1 + 2 * \sum_{k=1}^{n-1} \left(1 - \frac{|k|}{n}\right) \hat{\rho}_k\right)}$$

$$Since\ \hat{\rho}_k = \hat{\rho}_{-k}$$

Example: 10-Year U.S. Treasury Bond Rate



$$\bar{X} \pm 1.96 * \sqrt{\frac{\hat{\sigma}^2}{n}\left(1 + 2 * \sum_{k=1}^{n-1}\left(1 - \frac{|k|}{n}\right)\hat{\rho}_k\right)}$$

```
# assume bond data is in df called bond
MOE = 1.96*sqrt(vxbar)
LL = mean(bond$Adj.Close) - MOE
UL = mean(bond$Adj.Close) + MOE
```

We are 95% confidence that the mean bond rate is contained in the interval (2.21%, 2.45%).

```
> UL = m + 1.96*sqrt(vxbar)
> LL
[1] 2.214698961
> UL
[1] 2.449258338
```

DataScience@SMU

Estimating the Autocovariance

Review of Correlation

Recall: given a random sample from the bivariate distribution (X, Y)

X	Y
X_1	Y_1
X_2	Y_2
$\vdots$	$\vdots$
X_n	Y_n

$$r = \frac{\frac{1}{n}\,{}_{i=1}^{\;n}\left(X_i \quad \bar{X}\right)\left(Y_i \quad \bar{Y}\right)}{\sqrt{\frac{1}{n}\,{}_{i=1}^{\;n}\left(X_i \quad \bar{X}\right)^2}\sqrt{\frac{1}{n}\,{}_{i=1}^{\;n}\left(Y_i \quad \bar{Y}\right)^2}}$$

Average of all cross-product terms $(X_i \quad \bar{X})(Y_i \quad \bar{Y})$

Define Sample Autocorrelation (Lag 1)

X_t	X_{t+1}
X_1	X_2
X_2	X_3
:	:
X_{n-1}	X_n

"Average" of all cross-product terms $(X_t - \bar{X})(X_{t+1} - \bar{X})$

$$\hat{\rho}_1 = \frac{\frac{1}{n}\sum_{t=1}^{n-1}(X_t - \bar{X})(X_{t+1} - \bar{X})}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(X_t - \bar{X})^2}\sqrt{\frac{1}{n}\sum_{t=1}^{n}(X_t - \bar{X})^2}}$$

$$\boxed{= \frac{\frac{1}{n}\sum_{t=1}^{n-1}(X_t - \bar{X})(X_{t+1} - \bar{X})}{\frac{1}{n}\sum_{t=1}^{n}(X_t - \bar{X})^2}}$$

Define Autocovariance and Autocorrelation (Lag 2)

X_t	X_{t+2}
X_1	X_3
X_2	X_4
$\vdots$	$\vdots$
X_{n-2}	X_n

"Average" of all cross-product terms $(X_t - \bar{X})(X_{t+2} - \bar{X})$

$$\boxed{\hat{\rho}_2 = \frac{\frac{1}{n}\sum_{t=1}^{n-2}(X_t - \bar{X})(X_{t+2} - \bar{X})}{\frac{1}{n}\sum_{t=1}^{n}(X_t - \bar{X})^2}}$$

Define Autocovariance and Autocorrelation (Lag *k*)

X_t	X_{t+k}
X_1	X_{1+k}
X_2	X_{2+k}
:	:
X_{n-k}	X_n

Numerator: "Average" of all cross-product terms $(X_t - \bar{X})(X_{t+k} - \bar{X})$

$$\hat{\rho}_2 = \frac{\frac{1}{n}\sum_{t=1}^{n-k}(X_t - \bar{X})(X_{t+k} - \bar{X})}{\frac{1}{n}\sum_{t=1}^{n}(X_t - \bar{X})^2}$$

DataScience@SMU

Properties of Autocovariance

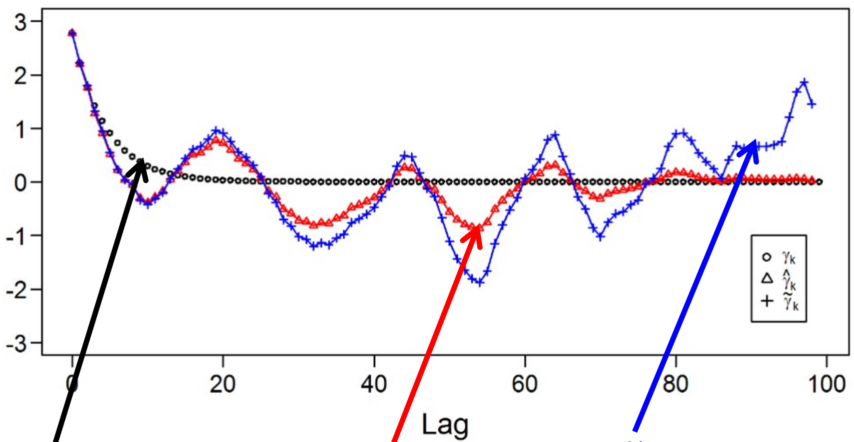
Definitions of Autocovariance

$1/n$ v $1/(n - k)$

Above and Beyond

Properties of $\hat{\ }_k$ (see text, pp. 18–22)

- Biased toward zero
- $\hat{\ }_k$ is less variable than $\tilde{\ }_k$ for k near n



~

True

^

(a) sample autocorrelations based on $\tilde{\ }_k$
do not satisfy properties (2) and (4)

DataScience@SMU

Stationary Time Series Wrap-Up

Key Point About Stationary Processes for Which $\rho_k \to 0$

For stationary time series, if the autocorrelations approach 0 as the distance between time points increases, then:

A single realization can be used to estimate:

- The mean
- The variance
- Autocorrelation
- The spectrum (covered next)

DataScience@SMU