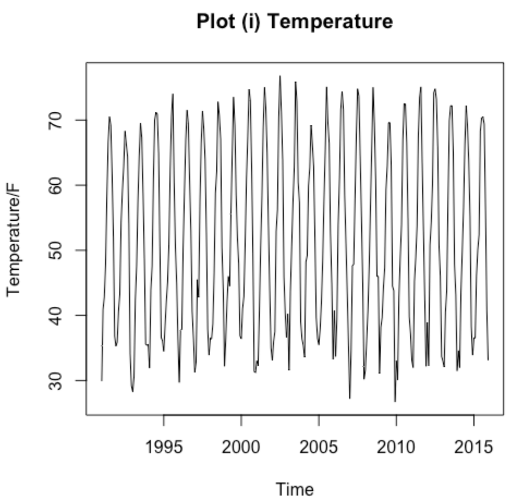
Solutions to Practice Midterm Exam

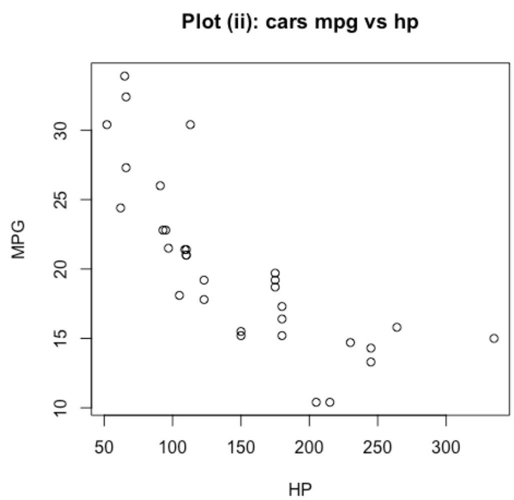
Q1. Which of the following data can be considered as time-series data?

a. daily sales of a grocery store
b. height of all the students in a college
c. final exam grades for students in a course
d. average salary of data analysts in Cincinnati

Time series data: a set of comparable measurements recorded on a single variable over multiple time periods.
Cross sectional data: measurements on multiple units, recorded in a single time period.
Panel data: cross-sectional measurements that are repeated over time, such as monthly health recordings for a sample of patients.

Q2. Which of the following plots is a time series plot?





a. Only (i)
b. Only (ii)
c. Both of them
d. Neither of them

Time series plot: X-axis should be time.

Q3. What is the Z-score corresponding to a 90% confidence level prediction interval?

1.645

Confidence Level	Z-score
90%	1.645
95%	1.960
99%	2.576

We can also use qnorm() function in R:

qnorm(0.95)

Q4. Suppose we have a series of observations as:

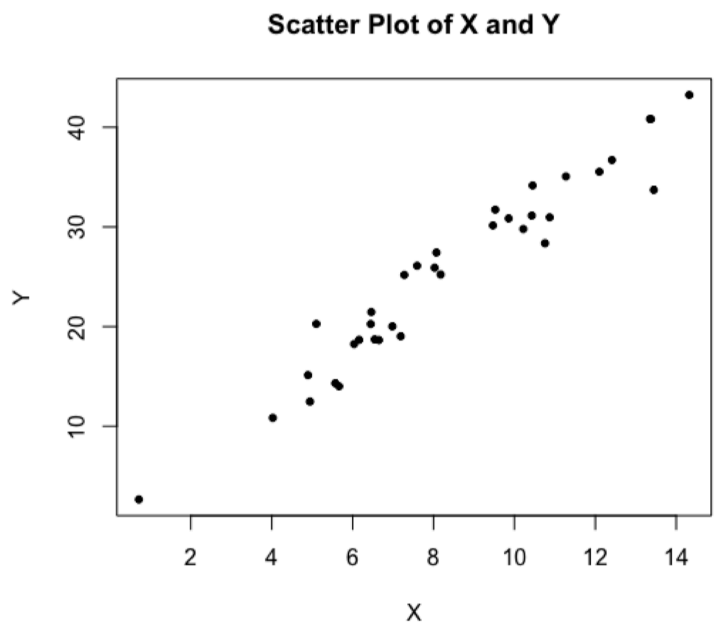
Time Period	Observation
1	8
2	15
3	14
4	13
5	13

What is the Z-score of the observation at time period 3? (Hint: you can use scale() function in R)

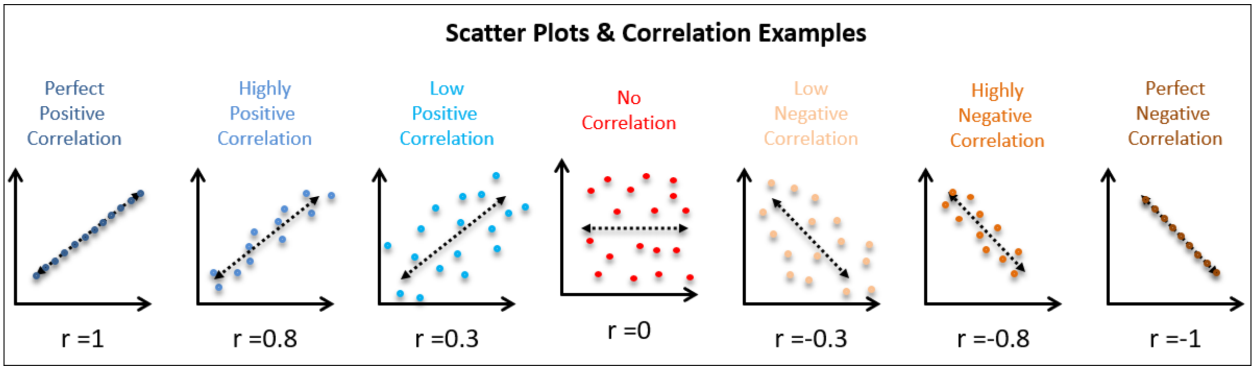
0.518

```
vec <- c(8, 15 , 14, 13, 13)
scale(vec)
```

Q5. Suppose we have a scatter plot of random variable X and random variable Y.



0.9



Q6. Suppose we have fitted a model for a time series data. We want to make a 95% percent confidence prediction interval. We have the point estimation as 1008.7, and we have the RMSE as 55.2. Which of the following is the correct prediction interval?

[900.5, 1116.9]

Lower bound: Point Estimation – Z * RMSE
Upper bound: Point Estimation + Z * RMSE

Z-score corresponding to 95%: 1.96
Lower bound: 1008.7 – 1.96 * 55.2 = 900.508
Upper bound: 1008.7 + 1.96 * 55.2 = 1116.892

Q7. Is the following statement correct?
"A model with lower mean error (ME) performs better in predicting."

Incorrect

A lower mean error can't guarantee a better performance of prediction.

Q8. Suppose we have fitted 2 models, Model 1 and Model 2, for a time series data. We have the following information for Model 1 and Model 2.

Model	Model 1	Model 2
In-sample MSE	753.6	890.1
Out-of-sample MSE	1120.4	922.5

Which of the model should we choose:

Model 2

A model with lower out-of-sample MAE/MSE/RMSE, performs better in predicting.

Q9. In$F_{t+h}(h)$, what does h stands for?

h: forecast horizon
t: forecast origin
t + h: time period we are predicting for

Q10. Suppose we have a time series data, and we are using a 3-order moving average model MA(3), and a 7-order moving average model MA(7). On the time series plot, which one is smoother?

MA(7)

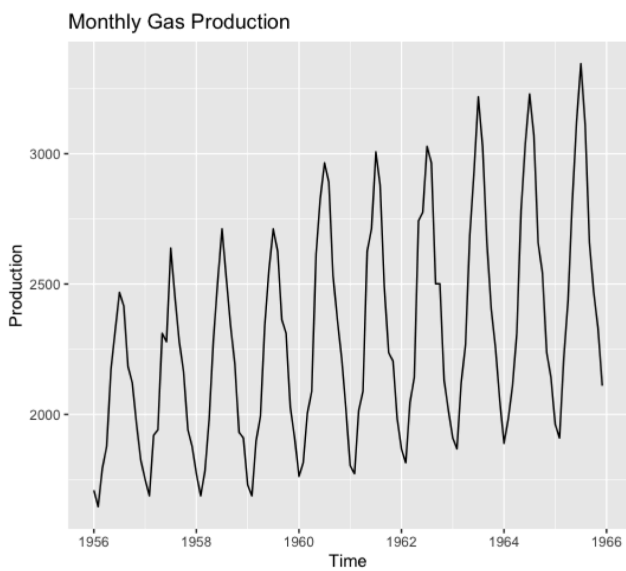
Larger value of k produces smoother plots, but is slower to adapt to changes.

Q11. Which of the following model can't include a local trend term?

Simple Exponential Smoothing

Simple Exponential Smoothing: level term
Linear Exponential Smoothing: level term, trend term
Holt-Winters Model: level term, trend term, seasonal components

Q12. Suppose we have a time series data of monthly gas production, and its time series plot is as the following:



Which of the following model should we consider?

Holt-Winters Model

From the time series plot, we can find there are seasonal factors. To include seasonal factors, we can use Holt-Winters Model.

Q13. For a Simple Exponential Smoothing model, which of the following statement is correct?

h-step ahead point forecast is equal to 1-step ahead forecast.

SES model can’t account local trend term.
A larger value of alpha gives a more rapid adjustment.
There are 1 smoothing parameter in SES, alpha.

Q14. Which of the following is a purely multiplicative model?

Y = T * S * E

Additive model: Y = T + S + E
Multiplicative model: Y = T*S*E
Mixed additive-multiplicative model: Y = T*S + E; or Y = (T + S)*E

Q15. Which one of the following is an observation equation?

A. $L_t = L_{t-1} + \alpha\epsilon_t$

B. $Y_t = L_{t-1} + T_{t-1} + \epsilon_t$

C. $L_t = L_{t-1} + T_{t-1} + \alpha\epsilon_t$

D. $T_t = T_{t-1} + \alpha\beta\epsilon_t$

Observation equation: An observation equation that relates the random variables (Yt, the actual value) to the underlying state variable(s).

Q16.

Which of the following is a state equation?

A. $Y_t = L_{t-1} + \epsilon_t$

B. $Y_t = L_{t-1} + T_{t-1} + \epsilon_t$

C. $L_t = L_{t-1} + T_{t-1} + \alpha\epsilon_t$

D. $Y_t = L_{t-1} + T_{t-1} + S_{t-m} + \epsilon_t$

State equations describe how the state variables evolve over time.

Q17. Suppose we have a sequence of monthly sales data, and we built an additive Holt-Winters Model. You find your estimated parameters are alpha = 0.2, beta = 0.45, gamma = 0.
Starting values of seasonal factors are 2.33, 12.15, -20.39, 5.91, and the starting value of the trend term is 25.44.
Which of the following statement is correct?

A. The model has fixed seasonal factors.
B. The model has no seasonal factors.
C. The model has a fixed trend.
D. The model has no trend term.

Special Cases:

- Fixed seasonal pattern: γ = 0 (no seasonal updating)
- No seasonal pattern: γ = 0 and all initial S values are set equal to zero
- Fixed trend: β = 0
- Zero trend: β = 0 and T_0 = 0
- All fixed components: $\alpha = \beta = \gamma = 0$

Q18. What are the state variables for a Holt-Winters Model?

L, T, S

State variables:
SES: L
LES: L, and T
Holt-Winters Model: L, T, and S

Q19. Which one of the following codes in ets() function represents a Mixed additive-Multiplicative Holt-Winters Model (Y = (L + T) * S * E)?

MAM

Method	Observation Equation	ETS
SES	Y = L + E	A,N,N
LES	Y = L + T + E	A,A,N
Multiplicative LES	Y = L * T * E	M,M,N
Seasonal SES	Y = S + E	A, N, A
Additive Holt-Winters	Y = L + T + S + E	A, A, A
Multiplicative Holt-Winters	Y = (L + T)* S * E	M, A, M

Q20. Suppose we have a sequence of time series data as following in the table. We want to use the centered moving average model of order 5, CMA(5), to make predictions. Which of the following should be the CMA(5) placed at t = 5?

Time Period	1	2	3	4	5	6	7	8
Value	59	49	58	42	52	49	49	54
CMA(5)			52.0	50.0	50.0	49.2		

(58 + 42 + 52 + 49 + 49)/5 = 50.0

Q21. Suppose we have a time series data, and we want to build a Linear Exponential Smoothing Model to make predictions.
The smoothing parameters we used are alpha = 0.5, beta = 0.2. Please fill in all the blanks in the following table.
Please round your results to 2 decimal places.
Hint: Please consider the following formulas:

- Update local level, in terms or one-step-ahead error:

$$L_t = L_{t-1} + T_{t-1} + \alpha e_t$$

- Update local trend after substituting for level:

$$T_t = T_{t-1} + \alpha\beta e_t$$

Time Period	Actual Values	Level	Trend	Forecast	Prediction Error
1	15.00				
2	17.00				
3	19.00				
4	20.00	19.00	2.00	[F_4]	-1.00
5	21.00	20.50	[T_5]	22.40	-1.40
6	22.00	21.70	1.76	23.46	-1.46
7	26.00	[L_7]	1.61	24.34	1.66
8	27.00	25.17	1.78	26.95	[E_8]

Forecast for time period 4:
Forecast = Level + Trend = 19.00 + 2.00 = 21.00
Or Forecast = Actual Values – Prediction Error = 20 – (-1.00) = 21.00

Trend for time period 5:
T_5 = T_4 + alpha * beta * Prediction_error_4 = 2.00 + 0.5 * 0.2 * (-1.00) = 1.90
Or T_5 = Forecast_5 – Level_5 = 22.40 – 20.50 = 1.90

Level for time period 7:
L_7 = Forecast_6 + alpha * Prediction_error_6 = 23.46 + 0.5 * (-1.46) = 22.73
Or L_7 = Forecast_7 – Trend_7 = 24.34 – 1.61 = 22.73

Prediction Error for time period 8:
E_8 = Actual_8 – Forecast_8 = 27.00 – 26.95 = 0.05

Q22.

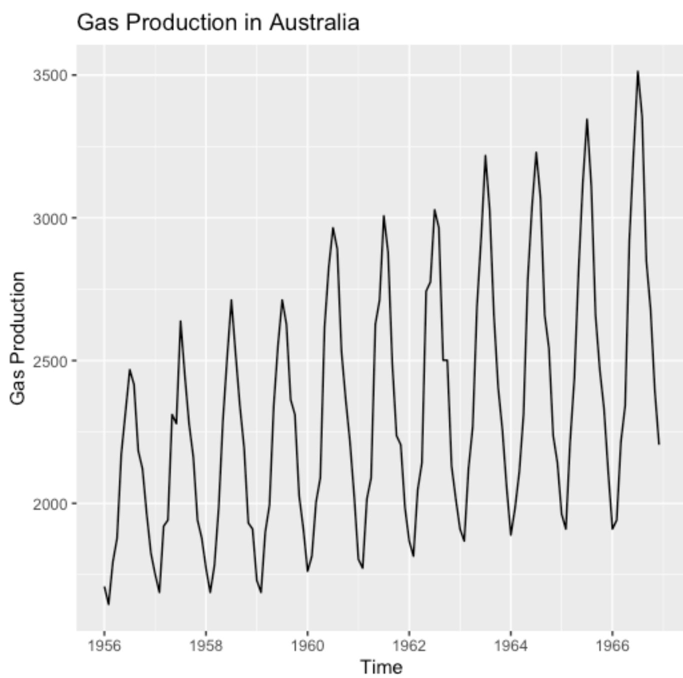
(1). Read in the data in R and create time series object:

```
library(forecast)
gas <- read.csv("Gas_Production.csv")
gas.ts <- ts(gas$Gas_Production, start = c(1956,1), frequency = 12)
```

For it is a monthly data, please remember to specify the frequency = 12.

(2). Please make a time series plot of the monthly gas production. Please comment on your time series plot. (10 pts).

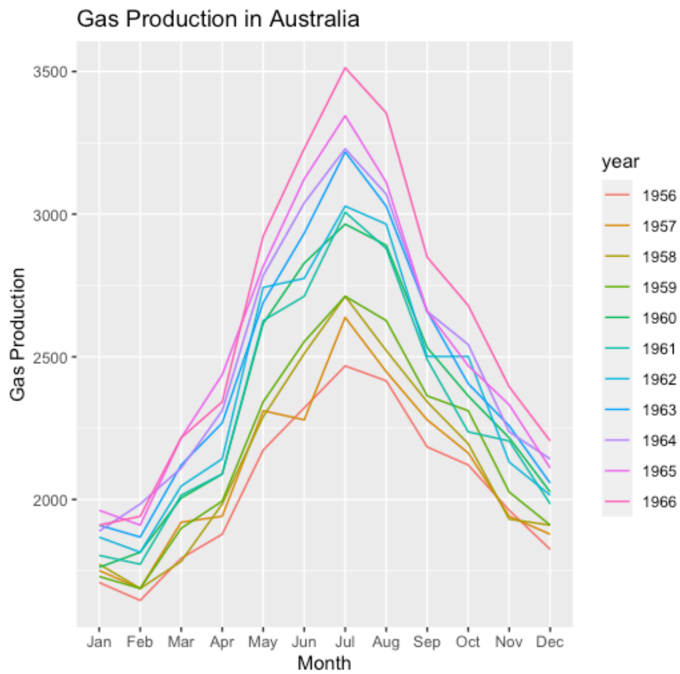
```
autoplot(gas.ts, main = "Gas Production in Australia", xlab = "Time", ylab = "Gas Production")
```



There is some seasonal pattern of gas production in Australia, and the amount of gas production increased over the time.

(3). Please make a seasonal plot of your data, and briefly comment on it. (5 pts)

ggseasonplot(gas.ts, main = "Gas Production in Australia", xlab = "Month", ylab = "Gas Production")



There is some seasonal pattern of the gas production. From February to July the gas production amount is increasing and, and it decreases from July to next year's February. Gas production achieves the annual maximum in July and annual minimum in February. This pattern repeats for every year.

(4). Please use the data from Jan 1956 to Dec 1965 to fit:
   (1). Additive Holt-Winters Model (AAA) (Y = L + T + S + E)
   (2). Holt-Winters model with multiplicative seasonal factor (MAM), Y = (L + T) * S * E (5 pts)

```
gas.10yrs.ts <- window(gas.ts, start = c(1956, 1), end = c(1965, 12))
model_add <- ets(gas.10yrs.ts, model = "AAA", damped = F)
model_mult <- ets(gas.10yrs.ts, model = "MAM", damped = F)
```

(5). Please use the Additive Holt-Winters model to construct a 95% confidence 1-stepahead prediction interval for Jan 1966 (5pts)

```
forecast(model_add, h = 1)
```

95% confidence interval: [1821.669, 2143.174]

(6). Please calculate the point estimation for all the months in 1966 using both model. (5pts)

```
frc_add <- forecast(model_add, h = 12)
frc_mult <- forecast(model_mult, h = 12)

frc_add$mean
frc_mult$mean
```

Prediction from additive model:

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1966	1982.421	1955.643	2162.259	2290.096	2732.473	2904.088	3096.994	2972.758	2659.061	2528.045	2312.680	2180.588

Prediction from multiplicative model:

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1966	2054.701	2026.692	2247.797	2386.107	2864.911	3055.831	3302.797	3146.490	2791.751	2638.452	2399.147	2239.277

(7). Please calculate the RMSE of the Additive Holt-Winters model and the Multiplicative Holt-Winters Model. Based on the RMSE you get, which model you would like to choose? (5 pts)

```
## Extract the actual values for year of 1966
actual_1966 <- window(gas.ts, start = c(1966,1), end = c(1966, 12))

## Calculate the Prediction errors
error_add <-   actual_1966 - frc_add$mean
error_mult <-   actual_1966 - frc_mult$mean

## Calculate the RMSE
MSE_add <- mean( error_add^2 )
MSE_mult <- mean( error_mult^2 )

## Calculate the RMSE
sqrt(MSE_add)
sqrt(MSE_mult)
```

RMSE of the additive model: 211.86, RMSE of multiplicative model: 114.91

I would choose the multiplicative model, for it has a lower out-of-sample RMSE, which indicates the multiplicative model performs better in predicting.