Module 3 Review Sheet

1. What is Autocorrelation Function (ACF)?

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

Autocorrelation Function (ACF) is a statistical correlation that measures the linear dependency of current observations with past observations or lags.

2. How is the ACF used in time series analysis?

The ACF is used to identify the appropriate MA (Moving Average) component in an ARIMA model. In an MA model, we would expect the ACF plot to show a sharp cut-off while the PACF plot displays a geometric or gradual decay.

3. What is Partial Autocorrelation Function (PACF)?

Partial Autocorrelation Function (PACF) is a measure of the correlation between observations at two time points, controlling for all other intervening time points. It provides the partial correlation of a time series with its own lagged values, controlling for the values of the time series at all shorter lags.

4. How is the PACF used in time series analysis?

The PACF is used to identify the appropriate AR (Auto-Regressive) component in an ARIMA model. In an AR model, we would expect the PACF to show a sharp cut-off and the ACF to display a geometric or gradual decay.

5. How to interpret the ACF and PACF plots?

Function	AR(*p*) Scheme	MA(*q*) Scheme	Mixed ARMA Scheme
ACF	Tails off as a damped wave pattern or damped exponential	Finite, *q* spikes	Tails off as a damped wave pattern or damped exponential
PACF	Finite, *p* spikes	Tails off as a damped wave pattern or damped exponential	Tails off as a damped wave pattern or damped exponential

6. How to use R to plot the ACF and PACF?

ggAcf() and ggPacf()

7. What is stationary?

A time series is stationary if it has a constant mean and variance, and its autocorrelations depend only on the relative time between the observations.

8. How can we test whether a time series data is stationary or not?

Augmented Dickey-Fuller Test (ADF test). ADF test is defined as:

H_0 (null hypothesis): the data series is not stationary
H_a (alternative hypothesis): the data series is stationary.

If we get a result of P-value greater than 0.05, then we can't reject the null hypothesis. The time series data tested is non-stationary.
If we get a result of P-value less than 0.05, then we can reject the null hypothesis. The time series data tested is stationary.

9. What does each component in ARIMA(p,d,q) stand for?

Autoregressive (AR)
Integrated (I)
Moving Average (MA)
p: the order of the autoregressive part
d: the degree of differencing involved.
q: the order of the moving average part

10. How is a Seasonal ARIMA model different from a regular ARIMA model?

Seasonal ARIMA contains 2 parts: ARIMA model for seasonal components, and ARIMA model for non-seasonal components.

11. What does each component in Seasonal ARIMA(p,d,q)(P,D,Q)m stand for?

p, d, q: as defined in Q9, for the non-seasonal part
P: the order of the seasonal autoregressive part
D: the degree of seasonal differencing
Q: the order of the seasonal moving average part
m: the number of periods in each season

12. How do you determine the order of differencing (d) in an ARIMA model?

The time series data is stationary after d-order differencing transformation.

13. What is information criteria? How to use it to select models?

Information criteria such as AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) are used to measure the goodness of fit of a statistical model. They are based on the concept of entropy from information theory. The idea is to find a balance between the complexity of a model and its goodness of fit to the data.

$$AIC = \ln(MSE) + 2p / n$$

$$BIC = \ln(MSE) + p\ln(n) / n$$

p = NP = number of AR parameters + number of MA parameters + 1 (for the variance) + 1 (if a constant is included)
A model with smallest AIC/BIC is usually the best model.

14. How to use R to fit an ARIMA or Seasonal ARIMA model?

If we want to fit an ARIMA mode with a given structure: to fit ARIMA(p, d, q) for a time series object data.
model <- arima(data, order = c(p, d, q))
to fit a seasonal ARIMA model ARIMA(p1, d1, q1)(p2, d2, q2)[m]
model <- arima(data, order = c(p1, d1, q1), seasonal = list(order = c(p2, d2, q2), period = m))

We can also use auto.arima() to automatically fit a best ARIMA model.

15. How to use R to forecast future values using a fitted ARIMA or Seasonal ARIMA model?

Suppose we have fitted an ARIMA model, and we name it as “model”. We can use function forecast() to make predictions for the future.
1-step-ahead forecast: forecast(model, h = 1)
h-step-ahead forecast: forecast(model, h = h)