Review Sheet for Module 1

1. What is time-series data?

A time series is a set of comparable measurements recorded on a single variable over multiple time periods.

2. Key characteristics of time series data: trend, seasonality, cycle. What do they look like in time series plots?

Trend: A time series contains a trend if it shows systematic movements over an extended period.

Seasonality: A time series has a season component if it displays a recurrent pattern with a fixed and known duration.

Cycle: A time series has a cyclical component if it displays somewhat regular fluctuations about the trend, but those fluctuations have a periodicity of variable and unknown duration, usually longer than one year.

3. Key difference of seasonality and cycle

Seasonality has a fixed and known duration for recurrent pattern, cycle has an unknown duration.

4. Difference between cross sectional data, time series data, and panel data.

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

5. What is time series plots and season plots? How to draw them in R? autoplot() for time series plot. seasonplot() and ggseasonplot() for season plots.

seasonplot() and ggseasonplot().
Please refer to Lab 2A.

6. How to make scatter plots in R?

7. Understand different measures of average: mean, median, and order statistics. Know how to use mean(), median(), sort() to calculate them.

mean() for mean.
median() for median.
sort() for order statistics. argument "decreasing = F" for an increasing order.

8. Understand different measures of variation: range, MAD, variance, and standard deviation.

$$\text{Range} = Y_{(n)} - Y_{(1)}$$
$$= \text{Largest} - \text{Smallest}$$
$$d_i = Y_i - \bar{Y}$$
$$MAD = \frac{\sum |d_i|}{n}$$
$$S^2(\text{ Variance }) = \frac{\sum d_i^2}{n-1}$$
$$SD = \sqrt{S^2}$$

max() to get the largest value. min() for the smallest value.
var() for variance.
sd() for standard deviation.
MAD: to calculate the MAD of vector x.

```
d <- x - mean(x)
MAD <- mean( abs(d) )
```

9. What is Z-score? How to caculate Z-score?

$$Z_i = \frac{Y_i - \bar{Y}}{S}$$

10. How to calculate corresponding probabilty for Z-score?

pnorm(x): the probability that Z is less than x.
x = qnorm(p): calculate the value which pnorm(x) = p

11. What is correlation coefficients?

It measures linear relationships. r ranges over [-1, +1]
Positive r means there is a positive linear relationship between the two variables. One increasing, the other also increase.
Negative r means there is a negative linear relationship between the two variables. One increasing, the other decrease.

r = 1 or -1 means there is a perfect positive/negative linear relationship between the two variables.
r = 0 means there is no linear relationship, but doesn't guarantee there is no other relationship, for example, quadratic or polynomial relationship.

12. Transformation: differences, growth rate and log transformation.

13. Understand the terminology of prediction.

Forecast origin, forecast horizon, one-step-ahead forecast, h-step-ahead forecast
one-step forecast error, h-step-ahead forecast error.
Rolling origin forecast.

14. Measures of forecasting accuracy: Why we don't use mean error and mean percentage error?

Because the positive error and negative error will cancel each other and make us underestimate the prediction error.

15. How to calculate the mean absolute error (MAE), mean absolute percentage error (MAPE), mean square error (MSE) and root mean square error (RMSE).

16. How to compare models? Which one's performance of prediction is better?

A lower out-of-sample RMSE.

17. How to construct prediction intervals?

1. Based on normal distribution.
2. Based on empirical distribution.