• Terms to verify that you know from previous classes (we won’t be covering unless you ask)
  • Random Variable, Mean, Expected Value, Variance, Standard Deviation
  • Population
  • Sample, sample mean, sample variance, sample standard deviation
  • Normal, T, Chi-square, F
  • Correlation, regression, residuals, SSE (sum of squared error = sum of squared residuals)
  • MSE (SE of regression)

• Terms that you may or may not be familiar with from previous classes (which we will be covering as we need them)
  • Covariance
  • Conditional Distribution, Conditional mean, conditional variance
  • Likelihood, log-likelihood, maximum likelihood
  • AIC, BIC (SIC)

• Homework assignment (due Tue, Jan 12)

I. Some conceptual/theoretical problems

1. Provide brief definitions or explanations for each term in the above “verify that you know” list.
2. Provide a brief, plain English, explanation of at least two of the terms on the “terms that you may or may not be familiar with” list (an explanation of what the thing is or how it is used is sufficient).

II. Some conceptual/theoretical problems

1. If \( y_t = \beta_0 + \beta_1 t + \epsilon_t \) and the \( \epsilon_t \) are independent of each other mean zero random variables, show that a centered moving average of length \( k \), explain why a 1st difference (i.e. \( y_t - y_{t-1} \)) would have no trend. [Hint: \( E(\epsilon_t) = 0 \).]
2. If we have a linear trend plus seasonality model: \( y_t = \beta_0 + \beta_1 t + \beta_{Jan} I_{Jan,t} + \beta_{Feb} I_{Feb,t} + \ldots + \beta_{Nov} I_{Nov,t} + \epsilon_t \) and the \( \epsilon_t \) are independent of each other mean zero random variables and \( I_{Month,t} \) is a month indicator, show that a first difference would still have seasonality.
3. With the linear trend plus seasonality model, show that a seasonal difference (i.e. \( y_t - y_{t-12} \)) would have neither trend nor seasonality.
4. In the formulation of the model in part (2), what is \( \beta_0 \)? Use plain English.
5. In the formulation of the model in part (2), what is \( \beta_1 \)? Use plain English.
6. In the formulation of the model in part (2), what is \( \beta_0 + \beta_1 t \)? Use plain English.
III. Some data oriented problems

1. For Beer production (R code to read this in is provided):
   a. Plot time series and describe the dataset.
   b. Create ACF and describe.
   c. For each series explain why a linear trend estimate from a regression would be appropriate way to “detrend” the series.
   d. If there is evident seasonality or trend, remove it with an appropriate difference then create ACF and describe
   e. Plot instead the log_{10} (whether natural log or log base 10 doesn’t matter, but log base 10 makes more sense with such economic data as this) of the Beer production. Comment on whether the log transformation would be more appropriate for analysis. [Hint: the log10() function in R can be used.]

2. Repeat a-d for the Global Temperature dataset.

3. Sketch time series of hourly temperature data for four days in SLO in February during a warming trend (heading from below average temps to above average temps). Sketch ACF for such a series.

IV. Yet another conceptual problem.

1. If \( SE_{ACF(k)} \) is approximately \( n^{-1/2} \), for each of the ACF values in parts 1b and 2b above, add a margin of error around the sample ACFs you created to provide an approximate 95% CI for the underlying ACF value for the process being observed.

2. Explain why such an approximate 95% CI would not include zero only if the sample ACF is outside the blue range provided by R’s acf() function.