Lab 3

Remember: to get help on any command you can type `?command`. Write-up due Friday Jan 30th. Give enough R code to show how you got any numerical answers you give.

1. **Getting familiar with the t distribution**

   We've talked a bit about the t distribution in class, but it can be really helpful to get familiar with distributions like this using a computer to simulate from the distribution, and by plotting graphs. You can simulate random numbers from the t distribution using the command `rt`. Try plotting histograms and qqnorm plots for random samples of 1000 values, using different numbers of degrees of freedom. Are the samples more normal-like for large or small degrees of freedom? Which plots make it easier to detect the deviations from normality? How do the expectation and variance vary with the degrees of freedom?

2. **Constructing prediction intervals: comparison of naive approach and predict.lm**

   Take the weather data you considered last time. Choose two cities you didn't use to fit the line. For each city use yesterday's (maximum) temperature to predict today's. Also find a 90% prediction interval for today's temperature at each of the two cities using three different methods: a) the "naive" method assuming a normal distribution which ignores the fact that your estimates of a, b and σ are not entirely accurate; b) a modified method using a t distribution, which (as we saw in class) allows for the fact that σ is estimated; c) the R command `predict.lm`. Which intervals are narrowest, and which are widest? Explain in your own words why this makes sense.

   Now, instead of making a symettric 90% prediction interval, use methods a) and b) above to give a lower bound for today's temperature at a 90% level of confidence: ie a value A such that the probability that A is smaller than today's temperature is 90%.

   a)