Homework 9 STAT 425/525
Due Monday November 14

R code for this assignment will be provided. I will need to see the plots you are asked to look at below.

1. We looked at the Chorley-Ribble cancer data in class. For this problem we focus on the locations of the 57 larynx cancer cases.

   (a) Compute the mean nearest neighbor distance. Carry out a randomization test to see if this mean distance is different from that expected under CSR. Provide me with a histogram of the nearest neighbor distances and a randomization p-value based on the alternative of clustering. What is your conclusion and why? I provide you with some of the R code but want you to come up with the rest on your own. See page 11 in the spatial notes for an example involving the entire Chorley data set. You should be able to modify this for the larynx cases only. There is an error in that code. The last line of it should be

   ```
   sum(hbar.vec<=hbar)/1000
   ```

   Provide me with a copy of the code you used.

   (b) Plot the theoretical and empirical G functions. What do they suggest about the spatial distribution of the points and why?

   (c) Plot the simulation envelope. Does it provide addtional evidence to back up your conclusion in part (a)? Why or why not.

2. There is a dataset in the `spatstat` package containing locations of 584 longleaf pines in a plot in Georgia. The investigators expected the distribution to be clustered.

   (a) Provide me with a plot of the pine locations. The plot shows circles of different diameters associated with marks based on tree size - ignore those.

   (b) Test the hypothesis of clustering against the null hypothesis of CSR using `quadrat.test`. Discuss the results.

   (c) It is possible to "plot" the results. The R code will do this for you. You will see 3 numbers in each grid cell, the observed count, the expected count, and (below those two) the standardized residual. Based on this graphical summary where does it appear CSR breaks down and how?

   (d) Below is the frequency distribution of the number of trees in a sample of 100 quadrats each of radius 6 meters.

Trees per quadrat	0	1	2	3	4	≥ 5
Observed Count	34	33	17	7	3	6
Expected Count						

The data were pooled for counts ≥ 5 to meet the assumptions of the method. Carry out a goodness-of-fit test based on an assumption of CSR. Give the expected frequencies under CSR. Discuss the results, paying particular attention to where CSR breaks down (if it does) - be careful because CSR can break down under both clustering and regularity. You are expecting clustering - do the results support that expectation and why? The sample mean of the observed counts was 1.43.

3. STAT 525: Suppose we have a realization of a spatial point process of n event locations. The cumulative distribution function of H (the nearest event-event distance) is the G function.

   (a) Derive the G function for a two-dimensional homogeneous Poisson process.

   (b) Find the probability density function of H.

   (c) Give the mean and variance of H. Hint: Before you start evaluating a gnarly integral or two take a close look at the pdf and see if you cannot identify the family of distributions. If you can do that you can use this knowledge to find the mean and variance.

4. The time in days to the development of a tumor for rats exposed to a carcinogen follows a Weibull distribution parameterized as in the notes with $\alpha = 2$ and $\lambda = 0.001$.

   (a) What is the probability a rat will be tumor free at 30 days? 60 days?

   (b) Find the hazard rate of the of the time to tumor appearance at 30 days. 60 days.

   (c) Find the median time to tumor development.