Due Date: October 31st, 2019
Problems due: 1,2,4,6

1. Finish in class Worksheet given on October 22nd, 2019.
2. χ^2 *goodness of fit test for continuous random variables*
   (a) Generate 1000 observations from uniform $[0, 1]$ in `R`.
   (b) Compute Y_j, the number of observations in $[\frac{j-1}{10}, \frac{j}{10}]$ with $j \in \{0, 1, \ldots, 10\}$.
   (c) Perform a χ^2 goodness of fit test on the null hypothesis being $p_1 = p_2 = \ldots p_{10} = \frac{1}{10}$ verus the alternative hypothesis being $(p_1, p_2, \ldots, p_{10}) \neq (\frac{1}{10}, \ldots, \frac{1}{10})$ for 5% level of significance.

   Do the above when 10 is replaced by $25, 50$, and 100.
3. *Two sample test of means for paired data* Let $n \geq 1$, $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ betwo samples, that are paired. Then the test Statistic is given by:

$$T := \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_Y)}{\frac{S}{\sqrt{n}}} \qquad \text{where} \qquad S^2 = \frac{1}{n-1}(\sum_{i=1}^{n}(z_i - \bar{z})^2) \qquad \text{with } z_i = x_i - y_i$$

   with $T \sim t_{n-1}$
   (a) Write a function called `pairedttest` that performs the above test for paired data x and y.
   (b) Consider the `shoes` dataset from the package `MASS`. It contains shoe wear data. There is a list of two vectors, giving the wear of shoes of materials A and B for one foot each of ten boys.
      i. Using the function `pairedttest` perform the paired t-test on the above data to see if the two types of shoes have different mean wear amounts.
      ii. Verify the same using the inbuilt function `t.test(A,B, paired=TRUE)`.
      iii. If you were to use the command `t.test` without the `paired=TRUE` command then can you still perform the above test ?
4. Execute the following `R` commands.

```
> x = seq(-10,10,by=1/10)
> y = 3*x^2 - 0.0001*x +7
> plot(y~x,pch=16, cex=0.2, col="red")
> abline(lm(y~x), col="blue")
```

   (a) Explain the reasons behind the poor simple regression line in contrast to the clear relationship between x and y
   (b) Using the `optim` command can you find estimates for $\beta_0, \beta_1, \beta_2$ that minimize $\sum_{i=1}^{100}(y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2)$
5. Let $\{T_i : i \geq 1\}$ be Exponential (λ) random variables. Define $X := \max\{j : \sum_{k=1}^{j} T_k \leq 1\}$
   (a) Show that for $l \in \mathbb{N}$,

$$\frac{\lambda^l}{l-1!}\int_x^\infty z^{l-1}e^{-\lambda z}dz = \sum_{m=0}^{l-1} e^{-\lambda x}\frac{(\lambda x)^m}{m!}$$

   (b) Show that $n \geq 0$, $P(X > n) = P(\sum_{k=1}^{n+1} T_k \leq 1)$
   (c) Suppose we have a random number generator that produces a sample U from a Uniform$(0, 1)$. Describe a simulation procedure from the above two parts that will generate a sample from Poisson (λ) using U.
6. Write a function called `NBsim` to simulate a sample from NegativeBinomial (n, p) from a U generated by using the `runif` command.
7. Write a function called `Csim` to simulate a sample from the standard Cauchy distribution from a U generated by using the `runif` command.