Due Date: February 24th, 2022
Problems due: 1(e), 2(a), 2(b), 3(a), 3(b)

1. Write functions[1] using `function(){}` command in `R` that do each the following:
   (a) Find the levels of factor of a given vector.
   (b) Change the first level of a factor with another level of a given factor.
   (c) Create an ordered factor from data consisting of the names of months.
   (d) Concatenate two given factors in a single factor.
   (e) Convert a given vector of integers to an ordered factor.
   (f) Extract the five of the levels of factor created from a random sample from the LETTERS.
   (g) Create a factor corresponding to height in `women` data set, which contains height and weights for a sample of women.

2. (Tschebychev Inequality)
   (a) Find a random variable X with $\text{Range}(X) = \{-1, 0, 1\}$ such that
   $$P(\mid X - \mu \mid \geq 2\sigma) = \frac{1}{4},$$
   with $\mu = E[X]$ and $\sigma^2 = \text{Var}[X]$.
   (b) Construct another random variable Y (different from X) with Range $(Y)= \{y_1, y_2, y_3\}$, mean μ and with
   $$P(\mid Y - \mu \mid > 2\sigma) > P(\mid X - \mu \mid > 2\sigma),$$
   so as to get
   $$P(\mid Y - \mu \mid > 2\sigma) > \frac{1}{4}$$
   Decide whether Tschebychev Inequality is violated ?
   (c) Write an `R`-code that takes an input k, and constructs a random variable X with $\text{Range}(X) = \{-1, 0, 1\}$ such that
   $$P(\mid X - \mu \mid \geq k\sigma) = \frac{1}{k^2},$$
   with $\mu = E[X]$ and $\sigma^2 = \text{Var}[X]$. Further the `R`-code should construct a random variable Y (different from X) with Range $(Y)= \{y_1, y_2, y_3\}$, mean μ so that
   $$P(\mid Y - \mu \mid > k\sigma) > \frac{1}{k^2}$$
   and (using replications) verifies your conclusion about Tschebychev's inequality in (b). It should save the entire output using `write.csv` as a (suitably designed) csv file.

3. De Moivre's Central Limit Theorem states the following:
   Theorem: *Let $X_n \sim$ Binomial(n, p), then*
   $$\lim_{n\to\infty} \left| P\left(\frac{X_n - np}{\sqrt{np(1-p)}} \leq x\right) - \int_{\infty}^{x} \frac{\exp(-\frac{y^2}{2})}{\sqrt{2\pi}} dt \right| = 0.$$

[1] Decide appropriately what the input and outputs should be. Each item should be separate function that can be called.

(a) Using the rbinom generate 100 samples of Binomial$(20, 0.5)$ and plot the histogram of the data-set.

(b) Using the rnorm generate 100 samples of Normal(10, 5) and plot the histogram of the data set.

(c) Provide a plot that can be thought off as a computer proof of the Central Limit Theorem.

4. (Poisson Approximation)

   (a) Using the rbinom generate 100 samples of Binomial(2000,0.001), save it in a dataframe dfbinomial and plot the histogram of the data-set.

   (b) Using the rpois generate 100 samples of Poisson(2), save it in a dataframe dfnormal and plot the histogram of the data set

   Think of ways you can enhance the above exercise to come up with a computer proof of the Poisson Approximation from Probability theory class.

5. *Extra Credit*[2]*.:* The following result is a Berry-Eseen Type bound

   Theorem: *Let* $X_n \sim$ *Binomial*(n, p)*, then there exists* $C > 0$ *such that*

$$\sup_{x \in R} \left| P\left(\frac{X_n - np}{\sqrt{np(1-p)}} \leq x\right) - \int_{\infty}^{x} \frac{\exp(-\frac{y^2}{2})}{\sqrt{2\pi}} dt \right| \leq \frac{(p^2 + (1-p)^2)}{2\sqrt{np(1-p)}}$$

   We shall prove it by simulation by the below algorithm.

```
For x = -2,-1.9,-1,8,...0,...,1.9,2
    using inbuilt pnorm find z[x]:- pnorm(x)
Set p
For m = 1,50,100,150,...,1000
    For x = -2,-1.9,-1,8,...0,...,1.9,2
      1)Generate B: 1000 Samples of Binomial (m,p) using inbuilt rbinom function
        Compute SB:  (B-m*p)/((m*p(1-p))^(0.5))
      2)Compute y[x] : the proportion of samples  in SB less than equal to x
      3)Repeat steps 1) and 2) 100 times and compute average --  my[x] over each trial.
      4)Calculate   diff[m]=  max(abs(my[x]- z[x]))

For m = 1,50,100,150,...,1000
    Calculate  error(m)= [p^2+(1-p)^2]/[2*(m*p*(1-p))^0.5]

Plot diff and error.
```

   (a) See if result is verified by picture. Can you do anything additional programitically in R to verify the Theorem ?

   (b) Prove DeMoivre's Central Limit Theorem and try to see if you can prove the bounds provided in this Theorem as well.

[2] The question is not part of the course syllabus. If I get a complete solution to this question [From two students] along with H.W and it is written up properly, along with a clear report then I will bake bannana muffins for the class when you all return later this semester or early next semester.