The University of British Columbia

Irving K. Barber Faculty of Science

DATA 101

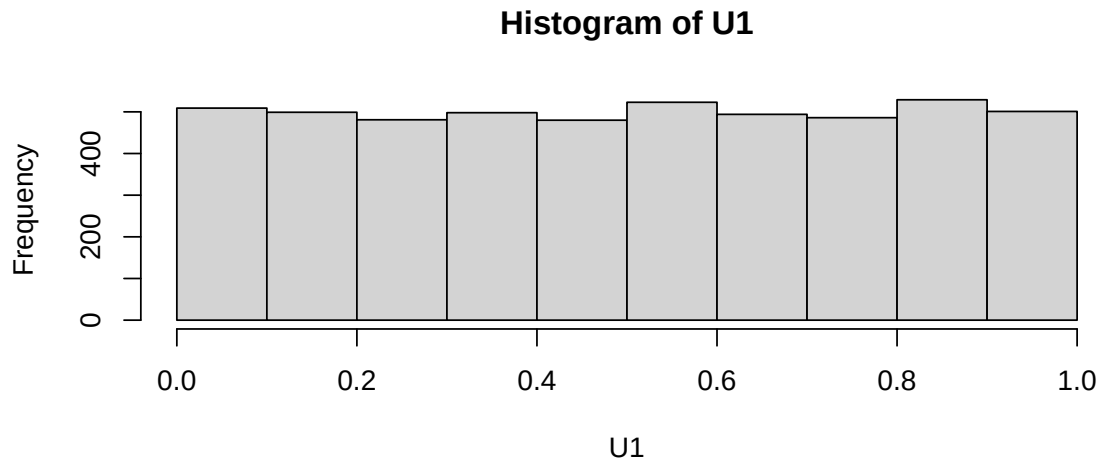
Lab 4 Solution

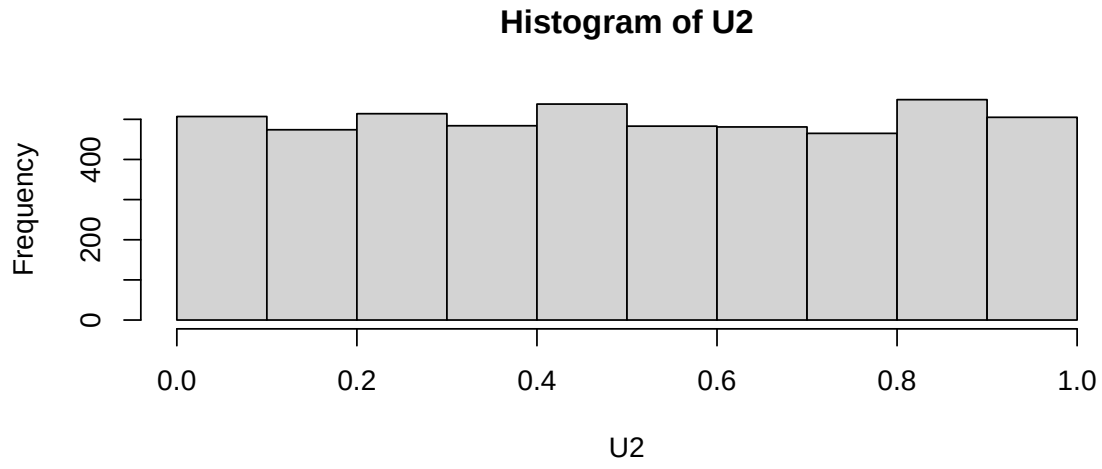
Date: October 5-9, 2020

Answer each question below, including the required lines of R code in each case.

1. Use the `runif()` function to simulate two vectors (`U1` and `U2`) of length 5000, each containing uniformly distributed random numbers. Plot histograms for the data in each vector and note the general shape of each histogram.

```
U1 <- runif(5000)
U2 <- runif(5000)
hist(U1); hist(U2)
```





The bars of the histograms are similar in height, giving a roughly rectangular shape.

2. Suppose you have pairs of uniform random numbers, and you always choose the smaller value, discarding the other one. If you plot a histogram of your collection of smaller values, what do you think it will look like? You can use the vectors `U1` and `U2` from the previous exercise to get an idea.

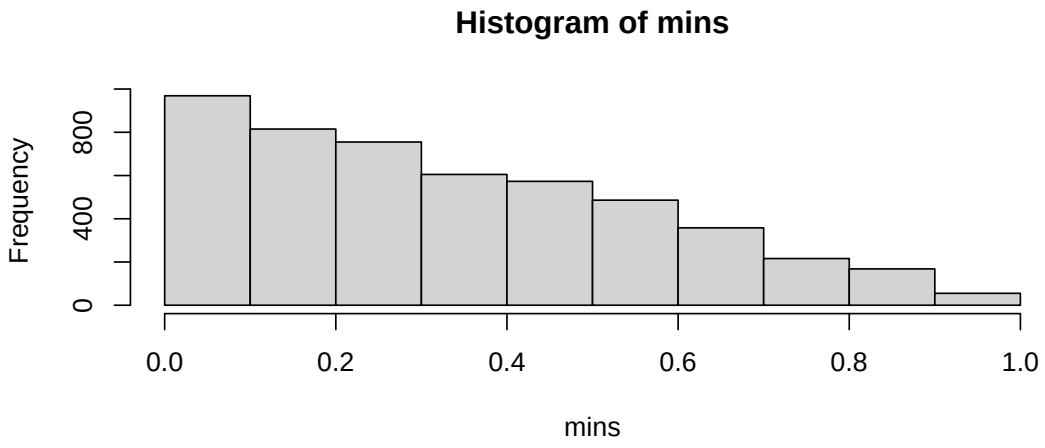
   You might expect smaller values to occur more often than larger values, since we will always be choosing the smaller value in each pair.

   (a) Use the `pmin()` function to obtain a vector `mins` of length 5000 which contains the pairwise minima of the vectors `U1` and `U2`.

```
mins <- pmin(U1, U2)
```

   (b) Plot the histogram of the pairwise minima, and note its shape.

```
hist(mins)
```



   The histogram now looks roughly like a triangle, decreasing in height from left to right.

3. Suppose you have pairs of uniform random numbers, and you always choose the larger value, discarding the other one. If you plot a histogram of your collection of larger values, what do you think it will look like?

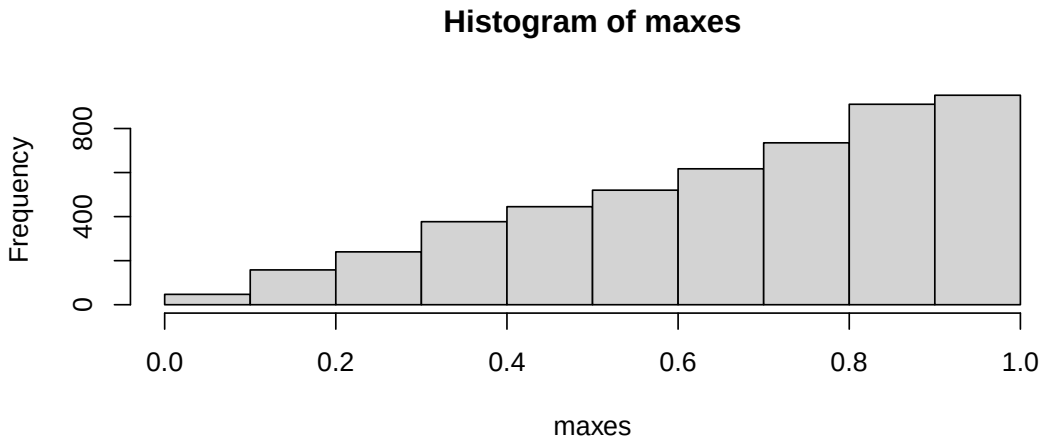
   You might expect larger values to occur more often than smaller values, since we will always be choosing the larger value in each pair.

   (a) Use the `pmax()` function to obtain a vector `maxes` of length 5000 which contains the pairwise maxima of the vectors `U1` and `U2`.

```
maxes <- pmax(U1, U2)
```

   (b) Plot the histogram of the pairwise maxima, and note its shape.

```
hist(maxes)
```



The histogram now looks roughly like a triangle, increasing in height from left to right.

4. Now, simulate another vector of uniform random values of length 5000 and store it in an object called `U3`.

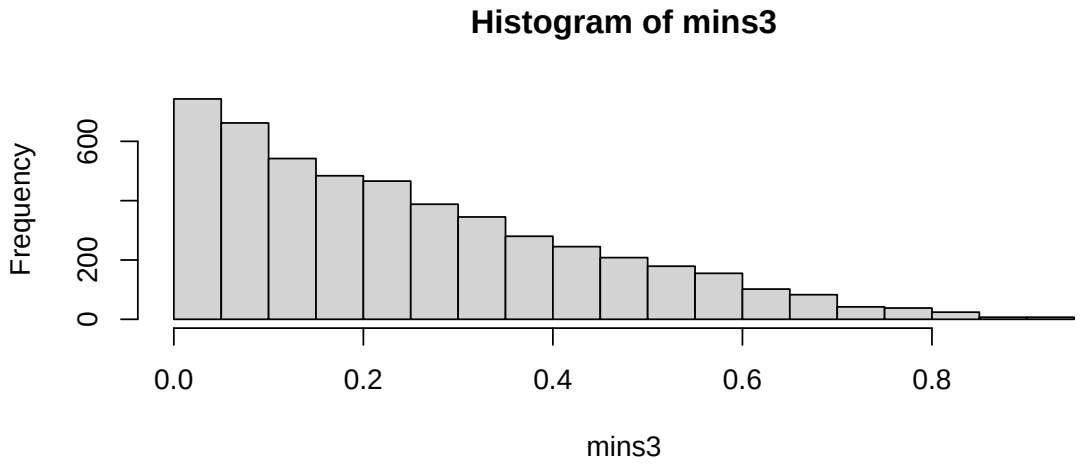
```
U3 <- runif(5000)
```

By applying the `pmin()` function twice, once to `U1` and `U2`, and then to `U3` and `mins`, we can obtain triplet-wise minima (i.e. the minimum of sets of 3 values). Having already obtained `mins`, we just need to execute the following code:

```
mins3 <- pmin(mins, U3)
```

Do this, and plot a histogram of the result, again noting the shape of the histogram.

```
hist(mins3)
```

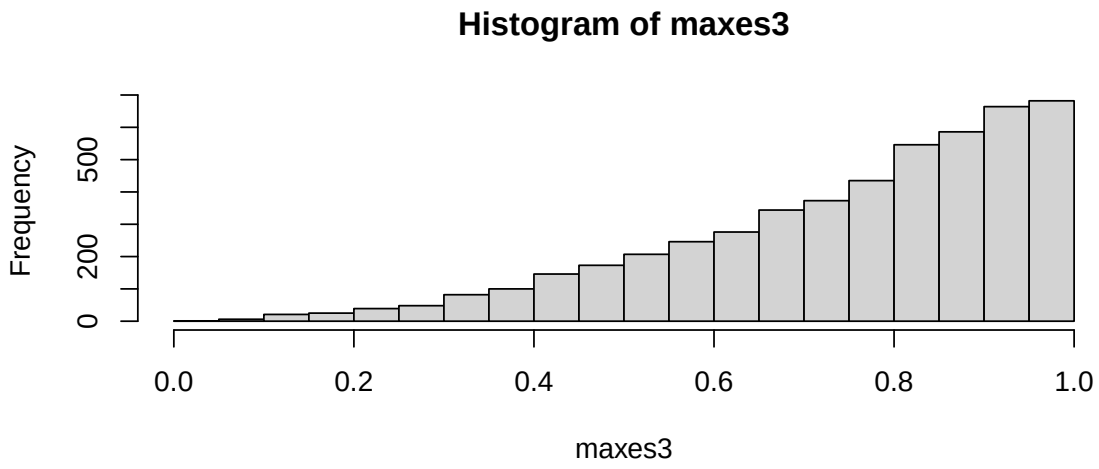


The histogram bars decrease in height from left to right, with the bars on the right side being shorter than for the pairwise minima.

5. Obtain triplet-wise maxima for the 5000 observations in `U1`, `U2`, and `U3`, assigning the results to a vector called `maxes3`. You may use `maxes` in your calculation. Plot the histogram and note its shape.

```
maxes3 <- pmax(maxes, U3)
```

```
hist(maxes3)
```



The histogram bars increase in height from left to right, with the bars on the left side being shorter than for the pairwise maxima.

6. If you have not already done so, install the *MPV* package either in RStudio or at the command line using

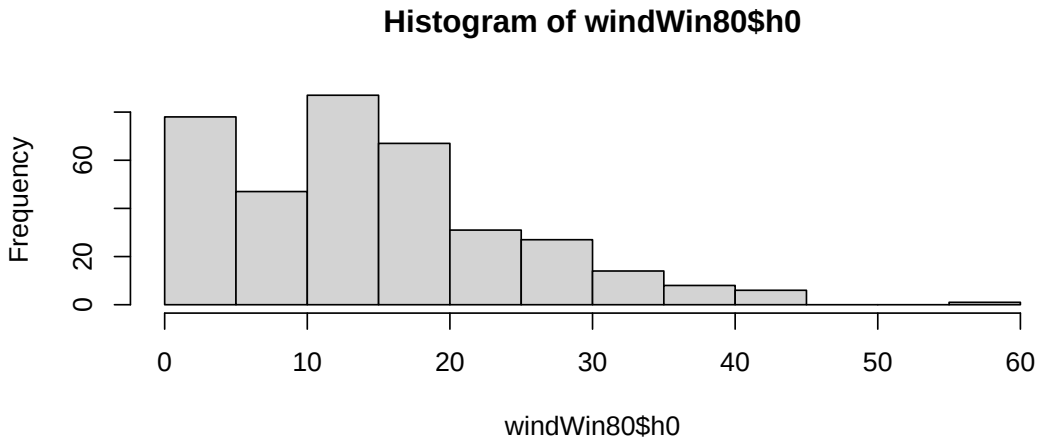
```
install.packages("MPV")
```

Load the package and consider the wind speed data in the data frame `windWin80`. The first column (`h0`) contains daily midnight wind speed measurements and the second column (`h12`) contains the wind speed measurements for the following noon hour. The measurements were recorded in 1980 at the Winnipeg International Airport.

```
library(MPV)
```

(a) Plot the histogram of the midnight wind speeds, and note the shape.

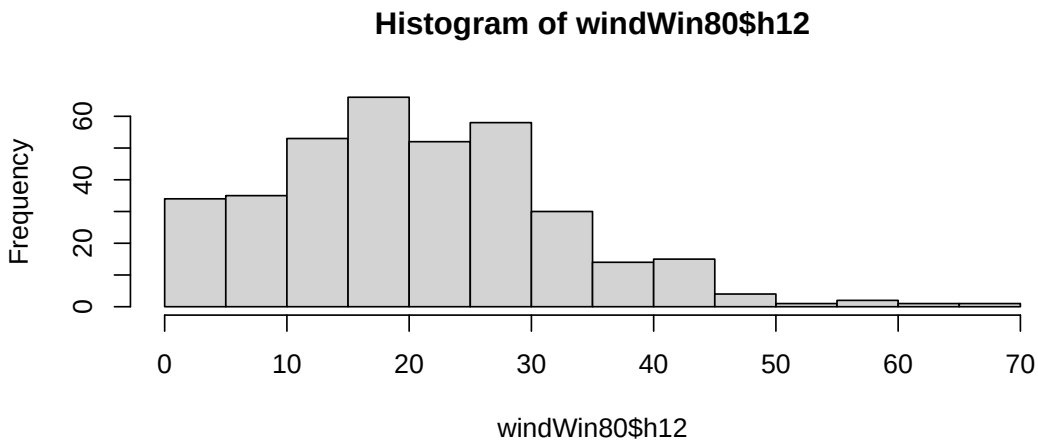
```
hist(windWin80$h0)
```



The histogram bars are tallest near 0 and decrease in height going from left to right.

(b) Plot the histogram of the noon hour wind speeds, and note the shape.

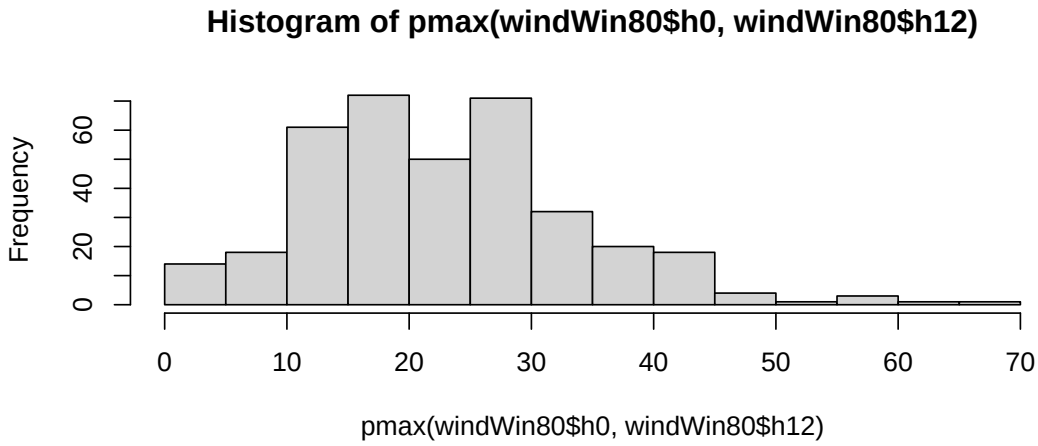
```
hist(windWin80$h12)
```



The histogram bars are tallest near 0 and decrease in height going from left to right.

(c) Plot the histogram of the pairwise maxima of the wind speeds, and note how the shape of the distribution has changed. (In climate studies, wind speed and rain are often studied by computing maxima and minima in this way; the climate scientists refer to the maxima as *block-maxima*.)

```
hist(pmax(windWin80$h0, windWin80$h12))
```



The histogram bars are tallest just to the left of the middle of the range. The shape of the distribution is more mound-shaped than for the separate noon hour and midnight wind speeds.

7. Continuing with the `windWin80` data frame, compare the midnight and noon hour wind speeds in the following ways. Specifically, for each of `h0` and `h12`, calculate the

   (a) means of the wind speeds. (Is it windier on average at noon or at midnight?)

```
mean(windWin80$h0); mean(windWin80$h12)
```

```
## [1] 14.59016
## [1] 21.14208
```

   It seems to be windier, on average, at noon.

   (b) medians of the wind speeds.

```
median(windWin80$h0); median(windWin80$h12)
```

```
## [1] 15
## [1] 20
```

   (c) standard deviations of the wind speeds. (Is the distribution of the wind speeds more spread out at noon or at midnight?)

```
sd(windWin80$h0); sd(windWin80$h12)
```

```
## [1] 10.44733
## [1] 12.04997
```

   The distribution is a little more spread out at noon than at midnight.

   (d) interquartile range of the wind speeds.

```
IQR(windWin80$h0); IQR(windWin80$h12)
```

```
## [1] 14
## [1] 15
```

(e) the maximum values of the wind speeds.

```
max(windWin80$h0); max(windWin80$h12)
```

```
## [1] 57
## [1] 70
```

Note that these are not the pairwise maxima, but rather the largest values recorded over the entire year.

8. Use the `windWin80` data frame to find the maximum noon hour wind speed when the wind speed for the previous midnight was less than 10.

```
max(windWin80$h12[windWin80$h0 < 10])
```

```
## [1] 43
```

9. Find the average of the noon hour wind speed values on those days where the wind speed at midnight had been between 5 and 25 (including both values).

```
Win5_25 <- subset(windWin80, h0 <= 25 & h0 >=5)
mean(Win5_25$h12)
```

```
## [1] 20.90517
```

10. Find the standard deviation of the noon hour wind speed values on those days where the wind speed at midnight had been between 10 and 30 (including neither of the values).

```
Win10_30 <- subset(windWin80, h0 > 10 & h0 < 30)
sd(Win10_30$h12)
```

```
## [1] 11.1412
```

11. Find the mean of the noon hour wind speed values on those days where either the noon hour wind speed is less than 50 or the midnight wind speeds are greater than the corresponding noon hour wind speeds.

```
Win50 <- subset(windWin80, h12 < 50 | h0 > h12)
mean(Win50$h12)
```

```
## [1] 20.525
```