Assignment 2

Re-submit Assignment

Due 6 Aug by 17:00   **Points** 20   **Submitting** a file upload

---

Assignment 2 – Due date: Aug 6, 5pm

Log in to Marvin.

In your home directory, create a directory called "Assignment_2".

Within Assignment_2, create two other directories called INPUT and OUTPUT.

Copy the file SWB_Full_garbled.txt.gz from /tutorials/Assignment_2/ to your INPUT directory.

This file contains the same information as the file SWB_Full.txt that you have worked with in Assignment 1.

SWB_Full.txt looked like this:

```
MarkerName        CHR      POS       A1        A2        EAF       Beta      SE        Pval
rs2075677         20       47701024  A         G         0.7743    0.021     0.004     1.879
rs238150          20       47851919  T         C         0.7743    0.022     0.004     2.012
rs6648   20       47895242         A         G         0.7761    0.022     0.004     2.579e-08
rs10485609        20       47665612  A         G         0.7743    0.021     0.004     2.954
rs17448715        20       47691743  T         C         0.222     -0.021    0.004     2.956
rs3021   20       47874005         A         G         0.7761    0.022     0.004     3.259e-08
rs348276          20       47786190  A         G         0.3209    -0.019    0.003     3.53e
rs1538482         20       47705496  T         C         0.6791    0.019     0.003     4.621
rs1885163         20       47701765  T         C         0.6791    0.018     0.003     5.262
```

SWB_Full_garbled.txt looks like:

```
ChrPosID rsID a1 a2 effect SE freq1 Pval
20:47701024  rs2075677 a g 0.021 0.004 0.7743 1.879e-
20:47851919  rs238150 t c 0.022 0.004 0.7743 2.012e-0
20:47895242  rs6648 a g 0.022 0.004 0.7761 2.579e-08
20:47665612  rs10485609 a g 0.021 0.004 0.7743 2.954e
20:47691743  rs17448715 t c -0.021 0.004 0.222 2.956e
20:47874005  rs3021 a g 0.022 0.004 0.7761 3.259e-08
20:47786190  rs348276 a g -0.019 0.003 0.3209 3.53e-0
20:47705496  rs1538482 t c 0.019 0.003 0.6791 4.621e-
20:47701765  rs1885163 t c 0.018 0.003 0.6791 5.262e-
```

In this assignment, we are asking you to convert SWB_Full_garbled.txt to SWB_Full.txt. Executing the commands asked by the following questions will get you there. Note down each command line you execute to implement the steps.

1. "SWB_Full_garbled.txt" is a space-delimited file, whereas SWB_Full.txt is tab-delimited. Use sed to change the delimiter in "SWB_Full_garbled.txt" from space to tab and write the output into a file named

"SWB_Full_garbled_tab.txt" in the OUTPUT directory.

2. You can do the same thing using awk. Write down the awk command that changes the delimiter of "SWB_Full_garbled.txt" from space to tab. (Don't execute this command since you already obtained the tab-delimited file using sed in the previous step).

3. Go into the OUTPUT directory. The first column in "SWB_Full_garbled.txt" contains information in the form CHR:POS, so you can obtain the CHR and POS columns (i.e. the 2nd and 3rd columns of "SWB_Full.txt") from the ChrPosID column. Using sed substitution instruction (**s**), change the colons (:) in the ChrPosID column to tabs. The output should have the same name. Use the relevant sed option for this (do not create another file and rename it "SWB_Full_garbled_tab.txt").

4. Now that you have separated the ChrPosID column into two, you have 9 columns in "SWB_Full_garbled_tab.txt", but the header (first line) only contains 8 column names. Change the ChrPosID string in the header to "CHR" and "POS" separated by tab to have 9 column names corresponding to the 9 columns.

5. The a1 and a2 columns contain the same information as the A1 and A2 columns in "SWB_Full.txt", but they are lower-case. Using awk, convert these columns to upper-case characters. Name the output file "SWB_Full_garbled_tab_UpCase.txt".

(Tip: awk has a built-in function that can do this.)

6. Now all the columns are in the file, but the column names and the order of the columns is different. Using awk, rename and reorder the columns so that the file looks exactly the same as "SWB_Full.txt" (freq1 = EAF, effect = Beta, rsID = MarkerName, the rest should be self-explanatory).

Bonus +2 points if you can do this in a single awk command!

7. Create another directory in "Assignment_2" directory named "CODE". Create a file named "YourName_txt" there using vim editor.

8. In vim editor, enter Input mode and write down your answers to steps 1-7, i.e. the command lines that you executed to implement steps 1-7. Save the file and quit.

You are done with the second assignment!

Upload the text file containing your answers ("YourName_assignment2.txt") to Canvas.