

The results below are generated from an R script.

```
#####  
# R-for-Epi  
# Epidemiologic Methods II (PHW250F, PHW250G)  
# created by Jade Benjamin-Chung  
  
# Solutions: Homework 2, Prevalence  
#####  
# Load okR autograder  
source('setup/autograder-setup/hw2_prev/hw2_prev.ok.R')  
  
## here() starts at /Users/Nolan/Desktop/grading-temp/hw-prev-trial copy  
## Warning: package 'checkr' was built under R version 3.5.2  
## Warning: package 'assertthat' was built under R version 3.5.2  
## Warning: package 'dplyr' was built under R version 3.5.2  
##  
## Attaching package: 'dplyr'  
## The following objects are masked from 'package:stats':  
##  
## filter, lag  
## The following objects are masked from 'package:base':  
##  
## intersect, setdiff, setequal, union  
  
AutograderInit()  
  
#####  
# Read in the data and view the data  
#####  
# Load the dplyr package  
library(dplyr)  
  
# This assignment uses data from the WASH Benefits  
# Bangladesh trial. The trial assessed whether  
# water, sanitation, handwashing, and nutrition  
# interventions delivered separately or together  
# could reduce child diarrhea and/or improve child  
# growth. The trials used a cluster- and block-randomized  
# design. Within each geographic block, 8 village clusters  
# were randomized to a treatment or control arm.  
# See Luby et al. 2018 for full details  
# (doi: http://dx.doi.org/10.1016/)  
  
# In this problem set we will calculate the prevalence  
# of diarrhea in different treatment arms after the  
# interventions were delivered.  
  
# The data and codebooks are publicly available here:  
# https://osf.io/pqzj5/  
  
# Load the diarrhea dataset:  
  
d = read.csv(paste0(here::here(), "/data/washb-data/washb-bangladesh-diar-public.csv"))
```

```

# Load the dataset with treatment variables:
tr = read.csv(paste0(here::here(), "/data/washb-data/washb-bangladesh-tr-public.csv"))

# Next let's merge the two datasets together.
# This will allow us to calculate the prevalence
# of diarrhea in different treatment arms
# (e.g., water, sanitation, handwashing, etc.)

d_tr = left_join(d, tr, by=c("block", "clusterid"))

# Now let's filter to only keep the rows for diarrhea
# measurements after the interventions were delivered.
# The svy variable includes values 0, 1, 2. We are going
# to drop the 0 values, which indicate the time period
# before interventions were delivered.
d_tr = d_tr %>% filter(svy!=0)

# Now we are going to drop children with missing values
# in the diarrhea variable from the dataset. This
# assumes that they were missing completely at random - i.e.,
# that there are no characteristics associated with whether
# a child was missing diarrhea measurement.
d_tr = d_tr %>% filter(!is.na(diar7d))

# Take a look at the merged dataset:
head(d_tr)

```

##	dataid	childid	tchild	clusterid	block	svy	month	sex	agedays	ageyrs
## 1	28001	C1	Sibling	280	1	2	10	female	1537	4.208076
## 2	28001	T1	Target child	280	1	2	10	male	659	1.804244
## 3	28001	T1	Target child	280	1	1	9	male	268	0.733744
## 4	28001	C1	Sibling	280	1	1	9	female	1146	3.137577
## 5	28002	C1	Sibling	280	1	1	9	female	1525	4.175222
## 6	28002	C1	Sibling	280	1	2	10	female	1916	5.245722

##	enrolage	newbirth	sibnewbirth	gt36mos	d3plus2d	d3plus7d	dloose2d	dloose7d	dblood2d
## 1	1.9329227	0	0	0	0	0	0	0	0
## 2	-0.4709103	1	0	0	0	0	0	0	0
## 3	-0.4709103	1	0	0	0	0	0	0	0
## 4	1.9329227	0	0	0	0	0	0	0	0
## 5	2.9705681	0	0	0	0	0	0	0	0
## 6	2.9705681	0	0	0	0	0	0	0	0

##	dblood7d	diar2d	diar7d	bruise2d	bruise7d	tooth2d	tooth7d	svyweek	svyyear	tr
## 1	0	0	0	0	0	0	0	42	2015	Sanitation
## 2	0	0	0	0	0	0	0	42	2015	Sanitation
## 3	0	0	0	0	0	0	0	38	2014	Sanitation
## 4	0	0	0	0	0	0	0	38	2014	Sanitation
## 5	0	0	0	0	0	0	0	38	2014	Sanitation
## 6	0	0	0	0	0	0	0	42	2015	Sanitation


```

# Problem 1: Calculate the number of children with
# diarrhea across all children in the dataset
# (ie., not stratifying by the treatment variable).
# Use the variable diar7d for diarrhea.

```

```

# Save your result in an object called p1.
# Label the result inside p1 as n_with_diarrhea
# Hint: Using the code from the tutorial, change
# the part that says "n_with_disease" to say
# "n_with_diarrhea"

```

```

p1 = d_tr %>%
  filter(diar7d==0) %>%
  summarise(x=n())

```

```
p1
```

```

##           x
## 1 15966

```

```
# Check your answer
```

```
CheckProblem1()
```

```
## Error: Did you remember to filter to only show results for children with diarrhea?
```

```

# Problem 2: Calculate the number of children without
# diarrhea across all children in the dataset
# (ie., not stratifying by the treatment variable).
# Use the variable diar7d for diarrhea.
# Save your result in an object called p2.
# Label the result inside p1 as n_without_diarrhea

```

```

p2 = d_tr %>%
  filter(diar7d==0) %>%
  summarise(n_without_diarrhea=n())

```

```
p2
```

```

## n_without_diarrhea
## 1                15966

```

```
# Check your answer
```

```
CheckProblem2()
```

```
## [1] "Correct!"
```

```
## Problem 2: 1/1
```

```

# Problem 3: Calculate diarrhea prevalence in the
# whole dataset (ignoring treatment arm) and
# save it in an object called prevalence.

```

```

prevalence = d_tr %>% summarise(prevalence=mean(diar7d))
prevalence

```

```

## prevalence
## 1 0.04549531

```

```
# Check your answer
```

```
CheckProblem3()
```

```
## [1] "Correct!"
## Problem 3: 1/1
```

```
# Problem 4: Now let's get counts of whether children
# did or did not have diarrhea in each treatment arm.
# In the tutorial, this created a 2x2 table for us.
# Here, since the WASH Benefits trial, there were
# 7 different arms (6 intervention + control)
# create a data frame with 14 rows (two for each arm).
# The first column is called "tr" for treatment.
# The second column is called "diar7d" and includes 0
# for children without diarrhea and 1 for children with
# diarrhea. The third column is called "n" and includes the
# number of children with or without diarrhea in that arm.

# Hint: the row for Control with no diarrhea should be this:
# tr          diar7d      n
# Control          0 3782
```

```
diar_tr_table = d_tr %>%
  group_by(tr,diar7d) %>%
  summarise(n=n())
```

```
diar_tr_table
```

```
## # A tibble: 14 x 3
## # Groups:   tr [7]
##   tr          diar7d      n
##   <fct>          <int> <int>
## 1 Control          0 3782
## 2 Control          1  240
## 3 Handwashing      0 1976
## 4 Handwashing      1   81
## 5 Nutrition        0 1974
## 6 Nutrition        1   78
## 7 Nutrition + WSH   0 2092
## 8 Nutrition + WSH   1   82
## 9 Sanitation       0 1995
## 10 Sanitation       1   73
## 11 Water           0 1998
## 12 Water           1  112
## 13 WSH             0 2149
## 14 WSH             1   95
```

```
# Check your answer
CheckProblem4()
```

```
## [1] "Correct!"
## Problem 4: 1/1
```

```
# Problem 5: Calculate the diarrhea prevalence in
# each treatment arm. You will need to combine
# different commands used in this problem set
# to calculate this. Save the results in an object
# called prevalence_tr. It should have 7 rows (one
# for each treatment) and two columns. The first
# column should be for the treatment name and the
# second should be for the prevalence.
```

```
prevalence_tr = d_tr %>%
  group_by(tr) %>%
  summarise(prevalence=mean(diar7d))
prevalence_tr

## # A tibble: 7 x 2
##   tr          prevalence
##   <fct>          <dbl>
## 1 Control        0.0597
## 2 Handwashing    0.0394
## 3 Nutrition      0.0380
## 4 Nutrition + WSH 0.0377
## 5 Sanitation     0.0353
## 6 Water          0.0531
## 7 WSH            0.0423

# Check your answer
CheckProblem5()

## [1] "Correct!"
## Problem 5: 1/1
```

```
# Problem 6: Examine the results in prevalence_tr.
# Which arm had the lowest diarrhea prevalence?
# Save the name of the treatment arm using the
# same spelling as in the treatment label in
# prevalence_tr to indicate your answer in an
# object called p6. (e.g., p6 = "Control")
```

```
p6 = "Sanitation"

# Check your answer
CheckProblem6()

## [1] "Correct!"
## Problem 6: 1/1
```

```
# Problem 7: Which treatment arm had prevalence
# closest to the prevalence in the control arm?
# Save the name of the treatment arm using the
# same spelling as in the treatment label in
# prevalence_tr to indicate your answer in an
# object called p6. (e.g., p7 = "Control")
```

```

p7 = "Water"

# Check your answer
CheckProblem7()

## [1] "Correct!"
## Problem 7: 1/1

# Check your total score
MyTotalScore()

##
## Problem 1: 0/1
## Problem 2: 1/1
## Problem 3: 1/1
## Problem 4: 1/1
## Problem 5: 1/1
## Problem 6: 1/1
## Problem 7: 1/1
## Total Score: 6/7

```

The R session information (including the OS info, R version and all packages used):

```

sessionInfo()

## R version 3.5.0 (2018-04-23)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Sierra 10.12.6
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] dplyr_0.8.3      assertthat_0.2.1 checkr_0.5.0    rlist_0.4.6.1   jsonlite_1.6
## [6] here_0.1         knitr_1.22
##
## loaded via a namespace (and not attached):
## [1] Rcpp_1.0.2      magrittr_1.5    tidyselect_0.2.5 R6_2.4.0
## [5] rlang_0.4.0     fansi_0.4.0     stringr_1.4.0   highr_0.7
## [9] tools_3.5.0     data.table_1.11.4 xfun_0.5        utf8_1.1.4
## [13] cli_1.0.1       rprojroot_1.3-2 tibble_2.0.1    crayon_1.3.4
## [17] purrr_0.3.1     glue_1.3.0      evaluate_0.14   stringi_1.3.1
## [21] compiler_3.5.0 err_0.2.0       pillar_1.3.1    backports_1.1.2
## [25] pkgconfig_2.0.2

Sys.time()

## [1] "2019-08-02 03:32:11 PDT"

```