Assignment #1 – Statistical software cheat sheet

**Download the dataset**

Depending on your disease of interest and preferred software, you should download:

|  |  |  |
| --- | --- | --- |
| **Software** | **Pneumonia** | **Diarrhea** |
| MS Excel | *Workshop\_BDA\_COI\_pneumonia.xlsx* | *Workshop\_BDA\_COI\_diarrhea.xlsx* |
| Stata 12+ | *Workshop\_BDA\_COI\_pneumonia.dta* | *Workshop\_BDA\_COI\_diarrhea.dta* |
| R | *Workshop\_BDA\_COI\_pneumonia.csv* | *Workshop\_BDA\_COI\_diarrhea.csv* |
| Any other | *Workshop\_BDA\_COI\_pneumonia.csv* | *Workshop\_BDA\_COI\_diarrhea.csv* |

As well as the *varname* workbook, which you will use to find your variables of interest and understand their structure.

|  |  |
| --- | --- |
| **Pneumonia** | **Diarrhea** |
| *Workshop\_BDA\_COI\_pneumonia\_varnames.xlsx* | *Workshop\_BDA\_COI\_diarrhea\_varnames.xlsx* |

**MS Excel**

Use pivot tables and apply filters for sector, type of visit, etc.

**Stata 12+**

***Parts 1 & 2: Household and government costs***

To generate a table of average costs by type of visit and sector:

by visit\_type facility\_sector, sort: ci …*COST VARIABLES…* (Stata 12-13) orby visit\_type facility\_sector, sort: ci means …*COST VARIABLES…* (Stata 14+)

Example:

by visit\_type facility\_sector, sort: ci consult\_1 diagnos\_1 medic\_1 bed\_1 food\_1 others\_1 consult\_2 diagnos\_2 medic\_2 bed\_2 travel\_2 food\_2 others\_2 consult\_3 diagnos\_3 medic\_3 bed\_3 travel\_3 food\_3 others\_3 total\_dmc total\_nmc total\_indirectcost total\_overallcost timelossdays\_main timelosshours\_main

***Part 3: Societal cost***

To calculate/generate a new variable (calculation provided as example):

gen NEW\_VARNAME = VARNAME\_1 + (VARNAME\_2 – VARNAME\_3) / 12

To replace missing observations (\*) with a value (here, zero):

replace NEW\_VARNAME = 0 if NEW\_VARNAME == \* (NEW\_VARNAME must exist already)

**R (R Studio)**

***Parts 1: Household costs***

To generate a table of average costs by type of visit and sector, you can complete and run the following procedure. The example provided for inpatient care in public facilities (grey dataset names can be changed):

require(tidyverse)

require(dplyr)

setwd(ENTER FOLDER PATH)

DATASET\_FULL <- read.csv("Workshop\_BDA\_COI\_pneumonia.csv")

DATASET\_Household <- select(DATASET\_FULL, facility\_sector, visit\_type, consult\_1, diagnos\_1, medic\_1, bed\_1, food\_1, others\_1, consult\_2, diagnos\_2, medic\_2, bed\_2, travel\_2, food\_2, others\_2, consult\_3, diagnos\_3, medic\_3, bed\_3, travel\_3, food\_3, others\_3, total\_dmc, total\_nmc, total\_indirectcost, total\_overallcost, timelossdays\_main, timelosshours\_main)

Household\_inpatient\_public <- filter(DATASET\_Household, visit\_type == "inpatient")

Household\_inpatient\_public <- filter(Household\_inpatient\_public, facility\_sector == "Public")

Household\_inpatient\_public$visit\_type <- NULL

Household\_inpatient\_public$facility\_sector <- NULL

TABLE\_Household\_inpatient\_public <- as.data.frame(sapply(Household\_inpatient\_public, mean, na.rm = TRUE))

view(TABLE\_Household\_inpatient\_public)

(Repeat for private for-profit sector and outpatient care)

***Part 2: Government cost***

To generate a table of average costs by type of visit, sector, and facility level, you can complete and run the following procedure. The example provided for inpatient care for facility level 2 (grey dataset names can be changed):

DATASET\_Government <- select(DATASET\_FULL, facility\_sector, facility\_type, facility\_level, visit\_type, facilitycost\_capital, facilitycost\_overhead, facilitycost\_labor, facilitycost\_supplies, facilitycost\_medications, facilitycost\_total)

Government\_inpatient\_lvl2 <- filter(DATASET\_Government, facility\_sector == "Public")

Government\_inpatient\_lvl2 <- filter(Government\_inpatient\_lvl2, visit\_type == "inpatient")

Government\_inpatient\_lvl2 <- filter(Government\_inpatient\_lvl2, facility\_level == "2")

Government\_inpatient\_lvl2$facility\_sector <- NULL

Government\_inpatient\_lvl2$facility\_type <- NULL

Government\_inpatient\_lvl2$visit\_type <- NULL

Government\_inpatient\_lvl2$facility\_level <- NULL

TABLE\_Government\_inpatient\_public <- as.data.frame(sapply(Government\_inpatient\_lvl2, mean, na.rm = TRUE))

view(TABLE\_Government\_inpatient\_public)

Note that the facility level 1 does not have any inpatient care.

***Part 3: Societal cost***

To generate a table of the different components of the societal cost by sector and type of visit, you can run the following procedure. You may need to install the *dplyr* R-package to use the *mutate* function. The example is for the societal costs attached to the use of public facilities for inpatient care (grey dataset names can be changed):

library(dplyr)

DATASET\_Societal <- select(DATASET\_FULL, facility\_sector, facility\_type, facility\_level, visit\_type, consult\_1, diagnos\_1, medic\_1, bed\_1, food\_1, others\_1, consult\_2, diagnos\_2, medic\_2, bed\_2, travel\_2, food\_2, others\_2, consult\_3, diagnos\_3, medic\_3, bed\_3, travel\_3, food\_3, others\_3, total\_dmc, total\_nmc, total\_indirectcost, total\_overallcost, timelossdays\_main, timelosshours\_main, facilitycost\_capital, facilitycost\_overhead, facilitycost\_labor, facilitycost\_supplies, facilitycost\_medications, facilitycost\_total)

DATASET\_Societal <- mutate(DATASET\_Societal, NetFacilityCost = case\_when(  
facility\_sector == "Public" ~ facilitycost\_total - consult\_2 - diagnos\_2 - bed\_2,  
facility\_sector != "Public" ~ 0))

DATASET\_Societal <- mutate(DATASET\_Societal, MedCostAtGovFacility = case\_when(  
facility\_sector == "Public" ~ consult\_2 + diagnos\_2 + bed\_2 + medic\_2,  
facility\_sector != "Public" ~ 0))

DATASET\_Societal <- mutate(DATASET\_Societal, RemainderMedCost = case\_when(  
facility\_sector == "Public" ~ total\_dmc - consult\_2 - diagnos\_2 - bed\_2 - medic\_2,  
facility\_sector != "Public" ~ total\_dmc))

DATASET\_Societal <- select(DATASET\_Societal, facility\_sector, facility\_type, facility\_level, visit\_type, NetFacilityCost, MedCostAtGovFacility, RemainderMedCost, total\_nmc, total\_indirectcost)

Societal\_inpatient\_public <- filter(DATASET\_Societal, facility\_sector == "Public")

Societal\_inpatient\_public <- filter(Societal\_inpatient\_public, visit\_type == "inpatient")

Societal\_inpatient\_public$facility\_sector <- NULL

Societal\_inpatient\_public$facility\_type <- NULL

Societal\_inpatient\_public$visit\_type <- NULL

Societal\_inpatient\_public$facility\_level <- NULL

TABLE\_Societal\_inpatient\_public <- as.data.frame(sapply(Societal\_inpatient\_public, mean, na.rm = TRUE))