obj3_all_tests.r

shivs

2026-01-08

```r
# obj 3. To analyze the impact of PMAGY on education and health facilities in the select
ed villages.

# import packages
library(readxl)
library(ltm)
```

```
## Loading required package: MASS
```

```
## Loading required package: msm
```

```
## Loading required package: polycor
```

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:MASS':
##
##     select
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(rcompanion)
# Load the purrr library for iteration
library(purrr)

post_pmagy_file_path <- "/home/shivs/r-analysis-pmagy/post_pmagy.xlsx"

### EDU Section tests:

print("EDU Section tests: ")
```

```
## [1] "EDU Section tests: "
```

```r
# POST PMAGY edu
edu_POST <- data.frame(read_excel(post_pmagy_file_path, sheet = "Sheet1", range = "AE1:A
N487")) # edu_POST
colnames(edu_POST) <- paste("POST", colnames(edu_POST), sep = "_")
# PRE PMAGY edu
edu_PRE <- data.frame(read_excel("/home/shivs/r-analysis-pmagy/pre_pmagy_obj3_edu.xlsx",
range = "A1:J487")) # edu_PRE
colnames(edu_PRE) <- paste("PRE", colnames(edu_PRE), sep = "_")



# calculate Cronbach's Alpha:
print(cronbach.alpha(edu_POST))
```

```
##
## Cronbach's alpha for the 'edu_POST' data-set
##
## Items: 10
## Sample units: 486
## alpha: 0.093
```

```r
print(cronbach.alpha(edu_PRE))
```

```
##
## Cronbach's alpha for the 'edu_PRE' data-set
##
## Items: 10
## Sample units: 486
## alpha: -0.07
```

```r
#

# Variance Inflation Factor (VIF):
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:purrr':
##
##     some
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
library(caTools)
# Fit a regression model on post data
edu_POST_w_index <- data.frame(read_excel(post_pmagy_file_path, sheet = "Sheet1", range
= "AE1:AO487"))
colnames(edu_POST_w_index) <- paste("POST", colnames(edu_POST_w_index), sep = "_")
model_all <- lm(POST_EDU_INDEX ~ ., data = edu_POST_w_index)
# Calculate VIF
vif_results <- car::vif(model_all)

print("VIF results for post data:")
```

```
## [1] "VIF results for post data:"
```

```
print(vif_results)
```

```
## POST_EDU.1....All.children.aged..6.17..years.in.your.household.are.attending.schoo
l...Primary..Middle..Secondary.and.Higher.Secondary.School.
##
1.020819
##                                        POST_EDU.2...All.children..18.23..in.your.household.
are.currently.attending.post.higher.secondary.education.
##
1.038576
##                                                         POST_EDU.3...Children.in.
our.household.receive.mid.day.meals.at.school.regularly.
##
1.091568
##                              POST_EDU.4...Children.in.my.household.have.benefited.fr
om.government.scholarships..Pre.metric.and.Post.Metric..
##
1.153481
##                                  POST_EDU.5...Availability.of.teachers.improved.
in.the.village.schools.since.the.implementation.of.PMAGY
##
1.134813
##                                                       POST_EDU.6...Childre
n.in.my.household.have.access.to.digital.learning.tools.
##
1.127888
##                                                     POST_EDU.7...Children.in.my.hou
sehold.actively.participate.in.online.education.programs
##
1.092862
##      POST_EDU.8...Has.anyone.in.your.family.ever.stopped.going.to.college.because.of.
difficulty.in.traveling.or.lack.of.transport.facilities.
##
1.019261
##          POST_EDU.9...Has.any.child.from.your.family.left.school.before.completing.i
t.due.to.financial.reasons.or.household.responsibilities.
##
1.015845
##                                                    POST_EDU.10....Students.i
n.our.household.attend.school.more.regularly.since.PMAGY.
##
1.023575
```

```
# Fit a regression model on pre data
edu_PRE_w_index <- data.frame(read_excel("/home/shivs/r-analysis-pmagy/pre_pmagy_obj3_ed
u.xlsx", range = "A1:K487"))
colnames(edu_PRE_w_index) <- paste("PRE", colnames(edu_PRE_w_index), sep = "_")

model_all <- lm(PRE_EDU_INDEX ~ ., data = edu_PRE_w_index)
# Calculate VIF
vif_results <- car::vif(model_all)
print("VIF results for pre data:")
```

```
## [1] "VIF results for pre data:"
```

```
print(vif_results)
```

```
## PRE_EDU.1....All.children.aged..6.17..years.in.your.household.are.attending.school...
Primary..Middle..Secondary.and.Higher.Secondary.School.
##
1.010827
##                                    PRE_EDU.2...All.children..18.23..in.your.household.a
re.currently.attending.post.higher.secondary.education.
##
1.030469
##                                        PRE_EDU.3...Children.in.o
ur.household.receive.mid.day.meals.at.school.regularly.
##
1.011279
##                                PRE_EDU.4...Children.in.my.household.have.benefited.fro
m.government.scholarships..Pre.metric.and.Post.Metric..
##
1.027301
##                                    PRE_EDU.5...Availability.of.teachers.improved.i
n.the.village.schools.since.the.implementation.of.PMAGY
##
1.017791
##                                            PRE_EDU.6...Childre
n.in.my.household.have.access.to.digital.learning.tools.
##
1.010098
##                                          PRE_EDU.7...Children.in.my.hous
ehold.actively.participate.in.online.education.programs
##
1.023414
##     PRE_EDU.8...Has.anyone.in.your.family.ever.stopped.going.to.college.because.of.d
ifficulty.in.traveling.or.lack.of.transport.facilities.
##
1.014436
##        PRE_EDU.9...Has.any.child.from.your.family.left.school.before.completing.it.
due.to.financial.reasons.or.household.responsibilities.
##
1.008338
##                                            PRE_EDU.10....Students.in.
our.household.attend.school.more.regularly.since.PMAGY.
##
1.011488
```

```
# TODO: Visualizing VIF Values from here https://www.geeksforgeeks.org/r-language/vif-fu
nction-in-r/
#

# Using theShapiro-Wilk Test :
print("Shapiro-Wilk Test results for pre data:")
```

```
## [1] "Shapiro-Wilk Test results for pre data:"
```

```
print(shapiro.test(edu_PRE_w_index$PRE_EDU_INDEX))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  edu_PRE_w_index$PRE_EDU_INDEX
## W = 0.98897, p-value = 0.001015
```

```
print("Shapiro-Wilk Test results for post data:")
```

```
## [1] "Shapiro-Wilk Test results for post data:"
```

```
print(shapiro.test(edu_POST_w_index$POST_EDU_INDEX))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  edu_POST_w_index$POST_EDU_INDEX
## W = 0.98917, p-value = 0.001174
```

```
#

# Paired Samples Wilcoxon Test :

# combine two data frames horizontally
edu_pre_post_w_index_hcomb <- cbind(edu_PRE_w_index, edu_POST_w_index)

# # Define a function to perform paired Paired Samples Wilcoxon Test between pairs of co
lumns
paired_wilcoxon_test <- function(before, after) {
  wilcox.test(before, after, paired = TRUE)
}

# Apply paired Paired Samples Wilcoxon Test to each pair of pre and post columns
results_wilcoxon_test <- map2(
  edu_pre_post_w_index_hcomb[grepl("PRE", names(edu_pre_post_w_index_hcomb))],
  edu_pre_post_w_index_hcomb[grepl("POST", names(edu_pre_post_w_index_hcomb))],
  paired_wilcoxon_test
)


# # Define a function to perform paired Paired Samples Wilcoxon Z Test between pairs of
columns
paired_wilcoxon_Z_VALS <- function(before, after) {
  wilcoxonZ(before, after, paired = TRUE)
}

# Apply paired Paired Samples Wilcoxon Z Test to each pair of pre and post columns
results_wilcoxon_Z_VALS <- map2(
  edu_pre_post_w_index_hcomb[grepl("PRE", names(edu_pre_post_w_index_hcomb))],
  edu_pre_post_w_index_hcomb[grepl("POST", names(edu_pre_post_w_index_hcomb))],
  paired_wilcoxon_Z_VALS
)

# # Print the results
print("Print the results for column/question wise Paired Samples Wilcoxon Test")
```

```
## [1] "Print the results for column/question wise Paired Samples Wilcoxon Test"
```

```r
for (i in 1:length(results_wilcoxon_test)) {
  Z <- results_wilcoxon_Z_VALS[[i]]
  names(Z) <- "Z"
  print(Z)

  tN <- length(results_wilcoxon_test)
  names(tN) <- "tN"

  r <- abs(Z) / sqrt(tN)
  names(r) <- "r"

  print(r)


  print(results_wilcoxon_test[[i]])
}
```

```
##      Z
## -16.4
##         r
## 4.944786
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  before and after
## V = 3363, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
##
##      Z
## -15.7
##         r
## 4.733728
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  before and after
## V = 5137.5, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
##
##      Z
## -19.5
##         r
## 5.879471
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  before and after
## V = 22, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
##
##      Z
## -19.6
##         r
## 5.909622
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  before and after
## V = 6, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
##
##      Z
## -19.6
##         r
## 5.909622
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  before and after
```

```
## V = 10.5, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
##
##      Z
## -18.9
##          r
## 5.698564
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  before and after
## V = 115, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
##
##    Z
## -19
##          r
## 5.728716
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  before and after
## V = 100, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
##
##     Z
## 12.7
##          r
## 3.829194
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  before and after
## V = 68380, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
##
##     Z
## 12.7
##          r
## 3.829194
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  before and after
## V = 60662, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
##
##     Z
## -13.8
##          r
## 4.160857
##
```

```
##   Wilcoxon signed rank test with continuity correction
##
## data:  before and after
## V = 6965.5, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
##
##      Z
## -19.1
##          r
## 5.758867
##
##   Wilcoxon signed rank test with continuity correction
##
## data:  before and after
## V = 0, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```

```r
#


### EDU Section tests COMPLETE
print("EDU Section tests COMPLETE")
```

```
## [1] "EDU Section tests COMPLETE"
```

```r
### HN Section tests:

print("HN Section tests: ")
```

```
## [1] "HN Section tests: "
```

```r
# POST PMAGY hn
hn_POST <- data.frame(read_excel(post_pmagy_file_path, sheet = "Sheet1", range = "AP1:AX
487"))
colnames(hn_POST) <- paste("POST", colnames(hn_POST), sep = "_")
# PRE PMAGY hn
hn_PRE <- data.frame(read_excel("/home/shivs/r-analysis-pmagy/pre_pmagy_obj3_hn.xlsx", r
ange = "A1:I487"))
colnames(hn_PRE) <- paste("PRE", colnames(hn_PRE), sep = "_")


# calculate Cronbach's Alpha:
print(cronbach.alpha(hn_POST))
```

```
##
## Cronbach's alpha for the 'hn_POST' data-set
##
## Items: 9
## Sample units: 486
## alpha: 0.055
```

```
print(cronbach.alpha(hn_PRE))
```

```
##
## Cronbach's alpha for the 'hn_PRE' data-set
##
## Items: 9
## Sample units: 486
## alpha: -0.09
```

```
#

# Variance Inflation Factor (VIF):
library(car)
library(caTools)

# Fit a regression model on post data
hn_POST_w_index <- data.frame(read_excel(post_pmagy_file_path, sheet = "Sheet1", range =
"AP1:AY487"))
colnames(hn_POST_w_index) <- paste("POST", colnames(hn_POST_w_index), sep = "_")

model_all <- lm(POST_HN_INDEX ~ ., data = hn_POST_w_index)
# Calculate VIF
vif_results <- car::vif(model_all)

print("VIF results for post data:")
```

```
## [1] "VIF results for post data:"
```

```
print(vif_results)
```

```
## POST_HN.1...The.Anganwadi.Centre.in.your.village.is.functioning.properly.in.terms.of.
building.condition..staff.availability..and.services
##
1.032062
##                                               POST_HN.2...Children..0.5.years..in.your.househ
old.are.healthy.and.generally.free.from.any.illness.
##
1.015870
##                                                   POST_HN.3....Health.centres.i
n.the.village.are.safe.and.accessible.for.childbirth.
##
1.012298
##                                               POST_HN.4....Vaccination.service
s.for.children.under.1.are.available.in.the.village.
##
1.031608
##                                                       POST_HN.5....The.ne
arest.health.centre.is.easy.to.access.from.our.home.
##
1.008427
##                                                   POST_HN.7....Ambulance.s
ervices.in.the.village.are.available.and.functional.
##
1.035254
##                                           POST_HN.8...Pregnant.women.in.the.household.re
ceived.adequate.antenatal.check.ups.during.pregnancy
##
1.010563
##                                                 POST_HN.9..Proper.treatment.was.provid
ed.for.managing.chronic.illnesses.in.your.household.
##
1.010637
##               POST_HN.10....Your.household.receives.proper.treatment.for.communicable.
diseases.such.as.malaria..TB..or.rabies.when.needed.
##
1.022533
```

```
# Fit a regression model on pre data
hn_PRE_w_index <- data.frame(read_excel("/home/shivs/r-analysis-pmagy/pre_pmagy_obj3_hn.
xlsx", range = "A1:J487"))
colnames(hn_PRE_w_index) <- paste("PRE", colnames(hn_PRE_w_index), sep = "_")

model_all <- lm(PRE_HN_INDEX ~ ., data = hn_PRE_w_index)

# Calculate VIF
vif_results <- car::vif(model_all)
print("VIF results for pre data:")
```

```
## [1] "VIF results for pre data:"
```

```
print(vif_results)
```

```
## PRE_HN.1...The.Anganwadi.Centre.in.your.village.is.functioning.properly.in.terms.of.b
uilding.condition..staff.availability..and.services
##
1.001861
##                                        PRE_HN.2...Children..0.5.years..in.your.househo
ld.are.healthy.and.generally.free.from.any.illness.
##
1.014545
##                                            PRE_HN.3....Health.centres.in.
the.village.are.safe.and.accessible.for.childbirth.
##
1.012797
##                                            PRE_HN.4....Vaccination.service
s.for.children.under.1.are.available.in.the.village.
##
1.004573
##                                                  PRE_HN.5....The.nea
rest.health.centre.is.easy.to.access.from.our.home.
##
1.007904
##                                                PRE_HN.7....Ambulance.se
rvices.in.the.village.are.available.and.functional.
##
1.006535
##                                      PRE_HN.8...Pregnant.women.in.the.household.rec
eived.adequate.antenatal.check.ups.during.pregnancy
##
1.011829
##                                            PRE_HN.9..Proper.treatment.was.provide
d.for.managing.chronic.illnesses.in.your.household.
##
1.014567
##              PRE_HN.10....Your.household.receives.proper.treatment.for.communicable.d
iseases.such.as.malaria..TB..or.rabies.when.needed.
##
1.013445
```

```
# TODO: Visualizing VIF Values from here https://www.geeksforgeeks.org/r-language/vif-fu
nction-in-r/
#

# Using theShapiro-Wilk Test :
print("Shapiro-Wilk Test results for pre data:")
```

```
## [1] "Shapiro-Wilk Test results for pre data:"
```

```
print(shapiro.test(hn_PRE_w_index$PRE_HN_INDEX))
```

```
## 
##  Shapiro-Wilk normality test
## 
## data:  hn_PRE_w_index$PRE_HN_INDEX
## W = 0.98762, p-value = 0.0003853
```

```
print("Shapiro–Wilk Test results for post data:")
```

```
## [1] "Shapiro–Wilk Test results for post data:"
```

```
print(shapiro.test(hn_POST_w_index$POST_HN_INDEX))
```

```
## 
##  Shapiro-Wilk normality test
## 
## data:  hn_POST_w_index$POST_HN_INDEX
## W = 0.98019, p-value = 3.519e-06
```

```r
#

# Paired Samples Wilcoxon Test :

# combine two data frames horizontally
hn_pre_post_w_index_hcomb <- cbind(hn_PRE_w_index, hn_POST_w_index)

# # Define a function to perform paired Paired Samples Wilcoxon Test between pairs of co
lumns
paired_wilcoxon_test <- function(before, after) {
  wilcox.test(before, after, paired = TRUE)
}

# # Apply paired Paired Samples Wilcoxon Test to each pair of pre and post columns
results_wilcoxon_test <- map2(
  hn_pre_post_w_index_hcomb[grepl("PRE", names(hn_pre_post_w_index_hcomb))],
  hn_pre_post_w_index_hcomb[grepl("POST", names(hn_pre_post_w_index_hcomb))],
  paired_wilcoxon_test
)


# # Define a function to perform paired Paired Samples Wilcoxon Z Test between pairs of
columns
paired_wilcoxon_Z_VALS <- function(before, after) {
  wilcoxonZ(before, after, paired = TRUE)
}

# Apply paired Paired Samples Wilcoxon Z Test to each pair of pre and post columns
results_wilcoxon_Z_VALS <- map2(
  hn_pre_post_w_index_hcomb[grepl("PRE", names(hn_pre_post_w_index_hcomb))],
  hn_pre_post_w_index_hcomb[grepl("POST", names(hn_pre_post_w_index_hcomb))],
  paired_wilcoxon_Z_VALS
)

# # Print the results
print("Print the results for column/question wise Paired Samples Wilcoxon Test")
```

```
## [1] "Print the results for column/question wise Paired Samples Wilcoxon Test"
```

```r
for (i in 1:length(results_wilcoxon_test)) {
  Z <- results_wilcoxon_Z_VALS[[i]]
  names(Z) <- "Z"
  print(Z)

  tN <- length(results_wilcoxon_test)
  names(tN) <- "tN"

  r <- abs(Z) / sqrt(tN)
  names(r) <- "r"

  print(r)


  print(results_wilcoxon_test[[i]])
}
```

```
##      Z
## -10.3
##          r
## 3.257146
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  before and after
## V = 14168, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
##
##      Z
## -17.9
##          r
## 5.660477
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  before and after
## V = 441, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
##
##    Z
## -18
##          r
## 5.6921
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  before and after
## V = 1148, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
##
##      Z
## -15.1
##          r
## 4.775039
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  before and after
## V = 5223.5, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
##
##      Z
## -14.1
##          r
## 4.458812
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  before and after
```

```
## V = 9015.5, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
##
##      Z
## -19.1
##         r
## 6.03995
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  before and after
## V = 0, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
##
##      Z
## -14.5
##          r
## 4.585303
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  before and after
## V = 11788, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
##
##      Z
## -8.37
##          r
## 2.646826
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  before and after
## V = 15186, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
##
##      Z
## -11.4
##          r
## 3.604997
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  before and after
## V = 12426, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
##
##      Z
## -19.1
##         r
## 6.03995
##
```

```
##  Wilcoxon signed rank test with continuity correction
##
## data:  before and after
## V = 1, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```

```
#


### HN Section tests COMPLETE
print("HN Section tests COMPLETE")
```

```
## [1] "HN Section tests COMPLETE"
```