TP9 : Merge des bases welfare EHCVM 2019 et 2021

LAWA FOUMSOU Prosper

2025-03-29

1 Introduction

Ce document présente le processus de fusion (**merge**) des bases de données *welfare* issues de l'**EHCVM** pour les années **2018** et **2021**.

2 Chargement des bases et détection des incohérences

```
# chargement des deux bases
base2018 <- read_dta("../data/ehcvm_welfare_sen2018.dta")
base2021 <- read_dta("../data/ehcvm_welfare_sen2021.dta")
```

Essayons d'avoir une idée sur la structure des deux bases.

Vérifions les variables présentes dans la base **EHCVM 2018** et non dans la base **EHCVM 2021.**

```
setdiff(names(base2018), names(base2021))
```

```
[1] "halfab"
```

Vérifions les variables présentes dans la base **EHCVM 2021** et non dans la base **EHCVM 2018.**

```
setdiff(names(base2021), names(base2018))
```

```
 [1] "month"                "hethnie"              "halfa"
 [4] "halfa2"               "def_temp_prix2021m11" "def_temp_cpi"
 [7] "def_temp_adj"         "zali0"                "dtet"
[10] "monthly_cpi"          "cpi2017"              "icp2017"
[13] "dollars"
```

Nous constatons que toutes les variables présentent dans la base de l'année 2018 sont présentent dans celle de l'année 2021. Sauf la variable relative à l'alphabétisation qui a été renommée différement dans dans la base 2021. Cependant, nous remarquons qu'il y a des variables présentes dans la base 2021 mais non dans la base 2019. Pour ces variables, les colonnes pour les individus de l'année 2018 seront vides car l'information n'a pas été collectée.

Corrigeons le nom de la variable relative à l'alphabétisation pour une harmonisation dans les deux bases.

```
base2021 <- base2021 %>%
  rename(
    halfab=halfa
  )
```

2.1 Détection des variables ayant des variables ayant un problème de codification ou de labélisation

Détectons les variables qui n'ont pas été codifiées ou labélisée de la même façon dans les deux bases.

```
variable_label_diff <- c() #créer une liste vide
vars_communes <- intersect(colnames(base2018), colnames(base2021))

for (variable in vars_communes) { #parcourir les variables en communs dans les deu
```

```
  if(labelled::is.labelled(base2018[[variable]])){ #vérifier si la variable en 201

      value_label_2018 <- labelled::val_labels(base2018[[variable]]) #recupérer le

  }else{
      value_label_2018 <- NULL #Mettre vide dans le cas ou la variable en 2018 n'e
  }

  if(labelled::is.labelled(base2021[[variable]])){ #vérifier si la variable en 202
value_label_2021 <- labelled::val_labels(base2021[[variable]]) #recupérer les labe

  }else{

      value_label_2021 <- NULL #Mettre vide dans le cas ou la variable en 2021 n'e
  }

  if(!identical(value_label_2018, value_label_2021)){ #Vérifier si les labels de l

     variable_label_diff <- append(variable_label_diff,variable) #Si les labels di
     print(variable)
   }
}
```

```
[1] "zae"
[1] "hnation"
[1] "hdiploma"
[1] "hactiv7j"
[1] "hbranch"
[1] "hcsp"
```

3 Récodification et labellisation des variables à problème

Afin de réussir la fusion de nos deux bases, nous devons harmoniser les modalités de mêmes variables dans les deux bases.

3.1 Correction de la variable hnation

Affichons la variable dans les deux bases pour avoir une idée sur la codification.

```
# Visualisation des modalités et codification
expss::val_lab(base2018$hnation)
```

```
                Benin              Burkina Faso             Côte d'Ivoire
                    1                         2                         3
       Guinée Bissau                      Mali                     Niger
                    4                         5                         6
             Sénégal                      Togo                   Nigéria
                    7                         8                         9
        Autre CEDEAO            Autre Afrique Autre pays hors Afrique
                   10                        11                        12
```

```
# Visualisation des modalités et codification
expss::val_lab(base2021$hnation)
```

```
                Bénin              Burkina Faso                 Cape-vert
                    1                         2                         3
       Cote d'ivoire                    Gambie                     Ghana
                    4                         5                         6
              Guinee             Guinée Bissau                   Liberia
                    7                         8                         9
               Mali                      Niger                   Nigeria
                   10                        11                        12
             Sénégal             Serra-Leonne                      Togo
                   13                        14                        15
       Autre Afrique Autre pays hors Afrique
                   17                        18
```

Affichons la distribution de la variable dans les deux bases pour une bonne récodification.

Characteristic	**N = 7,156**[1]
Nationalite du CM	
Benin	0 (0%)
Burkina Faso	1 (<0.1%)
Côte d'Ivoire	0 (0%)
Guinée Bissau	18 (0.3%)
Mali	27 (0.4%)
Niger	3 (<0.1%)
Sénégal	7,023 (98%)
Togo	2 (<0.1%)
Nigéria	2 (<0.1%)
Autre CEDEAO	39 (0.5%)
Autre Afrique	37 (0.5%)
Autre pays hors Afrique	4 (<0.1%)
Valeurs manquantes	0

[1] n (%)

```
# Distribution en 2018
base2018 %>%
  mutate(across(everything(), ~ labelled::to_factor(.))) %>%
  select(hnation) %>%
  tbl_summary(missing = "always", #afficher les valeurs manquantes
              missing_text = "Valeurs manquantes")
```

```
# Distribution en 2021
base2021 %>%
  mutate(across(everything(), ~ labelled::to_factor(.))) %>%
  select(hnation) %>%
  tbl_summary(missing = "always", #afficher les valeurs manquantes
              missing_text = "Valeurs manquantes")
```

Nous remarquons que la variable relative à la nationalité à été plus détaillée lors de l'enquête de l'année 2021. Par exemple dans la base 2018, Gambie, Ghana,Guinée,Cap-vert et Libéria ont été régroupés en Autres pays de la Cedeao

Characteristic	**N = 7,120**[1]
Nationalite du CM	
Bénin	0 (0%)
Burkina Faso	0 (0%)
Cape-vert	0 (0%)
Cote d'ivoire	1 (<0.1%)
Gambie	2 (<0.1%)
Ghana	1 (<0.1%)
Guinee	39 (0.5%)
Guinée Bissau	8 (0.1%)
Liberia	0 (0%)
Mali	18 (0.3%)
Niger	2 (<0.1%)
Nigeria	1 (<0.1%)
Sénégal	7,038 (99%)
Serra-Leonne	0 (0%)
Togo	1 (<0.1%)
Autre Afrique	7 (<0.1%)
Autre pays hors Afrique	2 (<0.1%)
Valeurs manquantes	0

[1] n (%)

alors qu'ils ont été explicités lors de l'enquête 2021. Cela change la codification de cette variable. Pour contourner cela nous allons uniformiser la codification à celle de 2018 qui est plus adaptée simple et adaptée ici.

```
# Recodage
base2021 <- base2021 %>%
  mutate(hnation = case_when(
    hnation == 1  ~ 1,   # Bénin -> Benin
    hnation == 2  ~ 2,   # Burkina Faso -> Burkina Faso
    hnation == 4  ~ 3,   # Cote d'Ivoire -> Côte d'Ivoire
    hnation == 8  ~ 4,   # Guinée Bissau -> Guinée Bissau
    hnation == 10 ~ 5,   # Mali -> Mali
    hnation == 11 ~ 6,   # Niger -> Niger
    hnation == 13 ~ 7,   # Sénégal -> Sénégal
    hnation == 15 ~ 8,   # Togo -> Togo
    hnation == 12 ~ 9,   # Nigeria -> Nigéria
    hnation %in% c(3, 5, 6, 7, 14) ~ 10,
    # Cape-vert, Gambie, Ghana, Guinée, Sierra Leone -> Autre CEDEAO
    hnation == 17 ~ 11,  # Autre Afrique -> Autre Afrique
    hnation == 18 ~ 12,  # Autre pays hors Afrique -> Autre pays hors Afrique
    TRUE ~ NA_real_
  ))

base2021$hnation <- labelled(
  base2021$hnation,
  labels = c(
    "Benin" = 1, "Burkina Faso" = 2, "Côte d'Ivoire" = 3,
    "Guinée Bissau" = 4,"Mali" = 5, "Niger" = 6, "Sénégal" = 7,
    "Togo" = 8, "Nigéria" = 9,"Autre CEDEAO" = 10,
    "Autre Afrique" = 11, "Autre pays hors Afrique" = 12
  )
)
```

3.2 Correction de la variable zae

Pour la récodification des autres variables, nous allons plutôt harmoniser les modalités des variables de la base de 2018 en fonction de celles de 2021 car elle

est plus actuelle.

Affichons la variable dans les deux bases pour avoir une idée de la codification et de la labélisation.

```
# Visualisation des modalités et codification
expss::val_lab(base2018$zae)
```

```
NULL
```

```
# Visualisation des modalités et codification
expss::val_lab(base2021$zae)
```

```
                        Kédougou              Saint-Louis-Matam
                               1                              3
          Thies-Diourbel-Louga        Kaolack-Fatick-Kaffrine
                               5                              7
Ziguinchor-Tamba-Kolda-Sédhiou                         Dakar
                               9                             11
```

Affichons la distribution de la variable dans les deux bases pour une bonne récodification.

```
# Distribution en 2018
base2018 %>% labelled::to_factor() %>%
  select(zae) %>%
  tbl_summary(missing = "always", #afficher les valeurs manquantes
              missing_text = "Valeurs manquantes")
```

```
# Distribution en 2021
base2021 %>% labelled::to_factor() %>%
  select(zae) %>%
  tbl_summary(missing = "always", #afficher les valeurs manquantes
              missing_text = "Valeurs manquantes")
```

Characteristic	N = 7,156[1]
Zone agroecologique	
1	1,020 (14%)
2	912 (13%)
3	1,602 (22%)
4	1,414 (20%)
5	1,752 (24%)
6	456 (6.4%)
Valeurs manquantes	0

[1] n (%)

Characteristic	N = 7,120[1]
Zone agroecologique	
Kédougou	452 (6.3%)
Saint-Louis-Matam	911 (13%)
Thies-Diourbel-Louga	1,599 (22%)
Kaolack-Fatick-Kaffrine	1,413 (20%)
Ziguinchor-Tamba-Kolda-Sédhiou	1,740 (24%)
Dakar	1,005 (14%)
Valeurs manquantes	0

[1] n (%)

Remarquon que la variable zae n'est pas labélisée dans la base de 2018. CependansRécodifions la variable zae dans la base 2018 et labélisont là en fonction de la base de 2021.

```
# Recodage
base2018 <- base2018 %>%
  mutate(
    zae = case_when(
      zae == 6 ~ 1,
      zae == 2 ~ 3,
      zae == 3 ~ 5,
      zae == 4 ~ 7,
      zae == 5 ~ 9,
      zae == 1 ~ 11,
      TRUE ~ NA_real_
    )
  )
base2018$zae <- labelled(
  base2018$zae,
  labels = c(
    "Kédougou" = 1,
    "Saint-Louis-Matam" = 3,
    "Thies-Diourbel-Louga" = 5,
    "Kaolack-Fatick-Kaffrine" = 7,
    "Ziguinchor-Tamba-Kolda-Sédhiou" = 9,
    "Dakar" = 11
  )
)
```

3.3 Correction de la variable hdiploma

Affichons la variable dans les deux bases pour avoir une idée de la codification et de la labélisation.

```
expss::val_lab(base2018$hdiploma)
```

```
Aucun        CEP/CFEE        BEPC/BFEM              cap               bt
```

```
             0              1              2              3              4
           bac  DEUG, DUT, BTS        Licence       Maitrise Master/DEA/DESS
             5              6              7              8              9
Doctorat/Phd
            10
```

```
expss::val_lab(base2021$hdiploma)
```

```
          Aucun            cepe            bepc             cap              bt
              0              1              2              3              4
           bac  DEUG, DUT, BTS        Licence       Maitrise Master/DEA/DESS
             5              6              7              8              9
Doctorat/Phd
            10
```

Affichons la distribution de la variable dans les deux bases pour une bonne récodification.

```
# Distribution en 2018
base2018 %>% labelled::to_factor() %>%
  select(hdiploma) %>%
  tbl_summary(missing = "always", #afficher les valeurs manquantes
              missing_text = "Valeurs manquantes")
```

```
# Distribution en 2021
base2021 %>% labelled::to_factor() %>%
  select(hdiploma) %>%
  tbl_summary(missing = "always", #afficher les valeurs manquantes
              missing_text = "Valeurs manquantes")
```

Pour cette variable, remarquons que le seul problème réside au niveau des labels, pour la codification il n' y a pas de problème.

Characteristic	N = 7,156[1]
Diplome du CM	
Aucun	5,697 (80%)
CEP/CFEE	587 (8.2%)
BEPC/BFEM	359 (5.0%)
cap	52 (0.7%)
bt	16 (0.2%)
bac	154 (2.2%)
DEUG, DUT, BTS	49 (0.7%)
Licence	83 (1.2%)
Maitrise	63 (0.9%)
Master/DEA/DESS	65 (0.9%)
Doctorat/Phd	31 (0.4%)
Valeurs manquantes	0

[1] n (%)

Characteristic	N = 7,120[1]
Diplome du CM	
Aucun	5,772 (81%)
cepe	583 (8.2%)
bepc	317 (4.5%)
cap	39 (0.5%)
bt	7 (<0.1%)
bac	150 (2.1%)
DEUG, DUT, BTS	46 (0.6%)
Licence	101 (1.4%)
Maitrise	55 (0.8%)
Master/DEA/DESS	32 (0.4%)
Doctorat/Phd	18 (0.3%)
Valeurs manquantes	0

[1] n (%)

```
base2018$hdiploma <- labelled(
  base2018$hdiploma,
  labels = c(
            "Aucun"=0, "cepe"=1, "bepc"=2, "cap"=3,
            "bt"=4, "bac"=5, "DEUG, DUT, BTS"=6, "Licence"=7,
            "Maitrise"=8, "Master/DEA/DESS"=9,"Doctorat/Phd"=10
  )
)
```

3.4 Correction de la variable hactiv7j

Affichons la variable dans les deux bases pour avoir une idée de la codification et de la labélisation.

```
expss::val_lab(base2018$hactiv7j)
```

```
             Occupe             Chomeur TF cherchant emploi    TF cherchant pas
                  1                   2                   3                   4
            Inactif      Moins de 5 ans
                  5                   6
```

```
expss::val_lab(base2021$hactiv7j)
```

```
             Occupe TF cherchant emploi    TF cherchant pas             Chomeur
                  1                   2                   3                   4
            Inactif      Moins de 5 ans
                  5                   6
```

Affichons la distribution de la variable dans les deux bases pour une bonne récodification.

```
# Distribution en 2018
base2018 %>% labelled::to_factor() %>%
  select(hactiv7j) %>%
  tbl_summary(missing = "always", #afficher les valeurs manquantes
              missing_text = "Valeurs manquantes")
```

Characteristic	N = 7,156[1]
Activite 7 jours du CM	
Occupe	5,362 (75%)
Chomeur	44 (0.6%)
TF cherchant emploi	3 (<0.1%)
TF cherchant pas	60 (0.8%)
Inactif	1,687 (24%)
Moins de 5 ans	0 (0%)
Valeurs manquantes	0

[1] n (%)

Characteristic	N = 7,120[1]
Activite 7 jours du CM	
Occupe	5,178 (73%)
TF cherchant emploi	5 (<0.1%)
TF cherchant pas	62 (0.9%)
Chomeur	34 (0.5%)
Inactif	1,841 (26%)
Moins de 5 ans	0 (0%)
Valeurs manquantes	0

[1] n (%)

```
# Distribution en 2021
base2021 %>% labelled::to_factor() %>%
  select(hactiv7j) %>%
  tbl_summary(missing = "always", #afficher les valeurs manquantes
              missing_text = "Valeurs manquantes")
```

Récodifions la variable relative à l'activité du CM durant les 7 jours précédant l'enquête. Pour cette variable, les labels sont differents pour les deux années.

```
base2018 <- base2018 %>%
  mutate(hactiv7j = case_when(
    hactiv7j == 1 ~ 1,   # Occupé -> Occupé
```

```
    hactiv7j == 2 ~ 4,   # TF cherchant emploi -> TF cherchant pas
    hactiv7j == 3 ~ 2,   # TF cherchant pas -> Chômeur
    hactiv7j == 4 ~ 3,   # Chômeur -> TF cherchant emploi
    hactiv7j == 5 ~ 5,   # Inactif -> Inactif
    hactiv7j == 6 ~ 6,   # Moins de 5 ans -> Moins de 5 ans
    TRUE ~ NA_real_      # Autres valeurs -> NA
  ))

# Ajouter les labels
base2018$hactiv7j <- labelled(
  base2018$hactiv7j,
  labels = c(
    "Occupe" = 1, "TF cherchant emploi" = 2, "TF cherchant pas" = 3,
    "Chomeur" = 4, "Inactif" = 5, "Moins de 5 ans" = 6
  )
)
```

3.5 Correction de la variable hcsp

Affichons la variable dans les deux bases pour avoir une idée de la codification et de la labélisation.

```
expss::val_lab(base2018$hcsp)
```

```
                                Cadre supérieur
                                              1
                  Cadre moyen/agent de maîtrise
                                              2
                    Ouvrier ou employé qualifié
                                              3
                Ouvrier ou employé non qualifié
                                              4
                       Manœuvre, aide ménagère
                                              5
                 Stagiaire ou Apprenti rénuméré
                                              6
```

```
                          Stagiaire ou Apprenti non rénuméré
                                                           7
Travailleur familial contribuant à une entreprise familiale
                                                           8
                               Travailleur pour compte propre
                                                           9
                                                      Patron
                                                          10
```

```
expss::val_lab(base2021$hcsp)
```

```
                                              Cadre supérieur
                                                            1
                                   Cadre moyen/agent de maîtrise
                                                            2
                                     Ouvrier ou employé qualifié
                                                            3
                                  Ouvrier ou employé non qualifié
                                                            4
                                         Manœuvre, aide ménagère
                                                            5
                                   Stagiaire ou Apprenti rénuméré
                                                            6
                               Stagiaire ou Apprenti non rénuméré
                                                            7
Travailleur Familial contribuant pour une entreprise familial
                                                            8
                                   Travailleur pour compte propre
                                                            9
                                                       Patron
                                                           10
```

Affichons la distribution de la variable dans les deux bases pour une bonne récodification.

Characteristic	N = 7,156[1]
CSP du CM	
Cadre supérieur	107 (2.0%)
Cadre moyen/agent de maîtrise	315 (5.8%)
Ouvrier ou employé qualifié	389 (7.2%)
Ouvrier ou employé non qualifié	490 (9.0%)
Manœuvre, aide ménagère	153 (2.8%)
Stagiaire ou Apprenti rénuméré	35 (0.6%)
Stagiaire ou Apprenti non rénuméré	11 (0.2%)
Travailleur familial contribuant à une entreprise familiale	100 (1.8%)
Travailleur pour compte propre	3,691 (68%)
Patron	143 (2.6%)
Valeurs manquantes	1,722

[1] n (%)

```
# Distribution en 2018
base2018 %>% labelled::to_factor() %>%
  select(hcsp) %>%
  tbl_summary(missing = "always", #afficher les valeurs manquantes
              missing_text = "Valeurs manquantes")
```

```
# Distribution en 2021
base2021 %>% labelled::to_factor() %>%
  select(hcsp) %>%
  tbl_summary(missing = "always", #afficher les valeurs manquantes
              missing_text = "Valeurs manquantes")
```

Pour cette variable, remarquons que le seul problème réside au niveau des labels, pour la codification il n' y a pas de problème.

```
base2018$hcsp <- labelled(
  base2018$hcsp,
  labels = c(
    "Cadre supérieur" = 1,
```

Characteristic	N = 7,120[1]
CSP du CM	
Cadre supérieur	57 (1.0%)
Cadre moyen/agent de maîtrise	280 (4.8%)
Ouvrier ou employé qualifié	450 (7.8%)
Ouvrier ou employé non qualifié	332 (5.7%)
Manœuvre, aide ménagère	151 (2.6%)
Stagiaire ou Apprenti rénuméré	34 (0.6%)
Stagiaire ou Apprenti non rénuméré	3 (<0.1%)
Travailleur Familial contribuant pour une entreprise familial	66 (1.1%)
Travailleur pour compte propre	4,302 (74%)
Patron	119 (2.1%)
Valeurs manquantes	1,326

[1] n (%)

```
    "Cadre moyen/agent de maîtrise" = 2,
    "Ouvrier ou employé qualifié" = 3,
    "Ouvrier ou employé non qualifié" = 4,
    "Manœuvre, aide ménagère" = 5,
    "Stagiaire ou Apprenti rénuméré" = 6,
    "Stagiaire ou Apprenti non rénuméré" = 7,
    "Travailleur Familial contribuant pour une entreprise familial" = 8,
    "Travailleur pour compte propre" = 9,
    "Patron" = 10
  )
)
```

3.6 Correction de la variable hbranch

Affichons la variable dans les deux bases pour avoir une idée de la codification et de la labélisation.

```
expss::val_lab(base2018$hbranch)
```

```
     Agriculture    Elevage/peche    Indust. extr.    Autr. indust.
```

```
             1              2              3              4
           btp       Commerce Restaurant/Hotel   Trans./Comm.
             5              6              7              8
Education/Sante Services perso.  Aut. services
             9             10             11
```

```
expss::val_lab(base2021$hbranch)
```

```
       Agriculture Elevage/syl./peche      Indust. extr.      Autr. indust.
                 1                  2                  3                  4
               btp           Commerce   Restaurant/Hotel       Trans./Comm.
                 5                  6                  7                  8
   Education/Sante    Services perso.      Aut. services
                 9                 10                 11
```

Affichons la distribution de la variable dans les deux bases pour une bonne récodification.

```
# Distribution en 2018
base2018 %>% labelled::to_factor() %>%
  select(hbranch) %>%
  tbl_summary(missing = "always", #afficher les valeurs manquantes
              missing_text = "Valeurs manquantes")
```

```
# Distribution en 2021
base2021 %>% labelled::to_factor() %>%
  select(hbranch) %>%
  tbl_summary(missing = "always", #afficher les valeurs manquantes
              missing_text = "Valeurs manquantes")
```

Pour cette variable, remarquons que le seul problème réside au niveau des labels, pour la codification il n' y a pas de problème.

Characteristic	N = 7,156[1]
Branche activite du CM	
Agriculture	1,366 (25%)
Elevage/peche	374 (6.9%)
Indust. extr.	58 (1.1%)
Autr. indust.	497 (9.1%)
btp	313 (5.8%)
Commerce	1,094 (20%)
Restaurant/Hotel	63 (1.2%)
Trans./Comm.	251 (4.6%)
Education/Sante	379 (7.0%)
Services perso.	761 (14%)
Aut. services	278 (5.1%)
Valeurs manquantes	1,722

[1] n (%)

Characteristic	N = 7,120[1]
Branche activite du CM	
Agriculture	1,363 (26%)
Elevage/syl./peche	523 (9.9%)
Indust. extr.	70 (1.3%)
Autr. indust.	643 (12%)
btp	361 (6.8%)
Commerce	904 (17%)
Restaurant/Hotel	75 (1.4%)
Trans./Comm.	296 (5.6%)
Education/Sante	421 (8.0%)
Services perso.	262 (5.0%)
Aut. services	354 (6.7%)
34	1 (<0.1%)
502	1 (<0.1%)
930	8 (0.2%)
Valeurs manquantes	1,838

[1] n (%)

```r
base2018$hbranch <- labelled(
  base2018$hbranch,
  labels = c(
    "Agriculture" = 1,
    "Elevage/syl./peche" = 2,
    "Indust. extr." = 3,
    "Autr. indust." = 4,
    "btp" = 5,
    "Commerce" = 6,
    "Restaurant/Hotel" = 7,
    "Trans./Comm." = 8,
    "Education/Sante" = 9,
    "Services perso." = 10,
    "Aut. services" = 11
  )
)
```

4 Jointure finale

Les labels étant harmonisés dans les deux bases, nous pouvons procéder à l'empilement des deux bases.

```r
# Ajouter les colonnes manquantes dans la base2018
col_manq_2018 <- setdiff(names(base2021), names(base2018))
base2018[col_manq_2018] <- NA

# Réordonner les colonnes pour être identiques
base2021 <- base2021[, names(base2018)]

# Fusionner
base_final <- dplyr::bind_rows(base2018, base2021)
```

Harmonisons les labels des variables dans la base finale.

```r
# Copier les labels de base2021 vers base_final
for (var in names(base2021)) {
  base_final[[var]] <- set_label(base_final[[var]], get_label(base2021[[var]]))
}
```

```r
dim(base_final)
```

```
[1] 14276    47
```

Enregistrement de la base finale

```r
write_dta(base_final, "../sorties/base_final.dta")
```