### **Problemas con Historical Data**

* ALB 1996 -> la llamamos EWS o LSMS como el resto de los años?

<https://microdata.worldbank.org/index.php/catalog/2311>

* BGD 1991 & 1995 -> nombre HHES, HES, o HIES? <https://catalog.ihsn.org/index.php/catalog/112>

<https://catalog.ihsn.org/index.php/catalog/152>

* BRA 1981-1985 -> los ginis dan MUY cercanos (**con cohh==1)** pero no exactos:

|  |  |  |
| --- | --- | --- |
|  | **Povcalnet** | PremCopied\Copied\_Oct2018 |
| 1981 | 0.579519 | .579309 |
| 1982 | 0.584055 | .584171 |
| 1983 | 0.589967 | .589880 |
| 1984 | 0.583686 | .583760 |
| 1985 | 0.555714 | .555868 |
| 1986 | 0.584646 | .584646 |
| 1987 | 0.596928 | .596928 |
| 1988 | 0.614284 | .614284 |
| 1989 | 0.632951 | .632951 |

* Eliminé los cohh = 0 al guardar en GMD
* CAF 2003 -> El Master dice que es ESCVM, Rural - Povcalnet/pnc\_pf dice que es ECASEB, Nacional
* EGY ->
* no hay datos de 1990
* las bases de 1995 y 1999 en P:\Documentation\PremCopied no replican el Gini
* cambié en el master chart la sigla de HIES a HIESC de 1990 a 2015 para ser consistente con los nombres en P:\01.PovcalNet\01.Vintage\_control\EGY, y con el último año, 2017
* GMB ->
  + 2003 -> Gini no replica.
  + La sigla para esta encuesta (Integrated Household Survey) debe ser IHS o HIS?. La sigla difiere en Master vs. Vintage folder en todos los años.
  + 2010 -> El master dice que es Household Poverty Survey (HPS) y Vintage folder que es Integrated Household Survey (IHS)
* IDN (Indonesia) ->
  + Faltan las bases de 1984, 1987 & 1990
  + Las de 1993, 1996, 1998 & 1999 están en P:\Documentation\PremCopied y replican el Gini Urbano y Rural, pero NO el Nacional. R// its ok. La distribución NAL es otra y se crea a parte
  + **Pregunta**: En el folder que creé en Vintage, guardamos 1 sola base \_HIST.dta? O una por área? R// 1 solo dta.
  + JAM ->
* No datos para 1988 & 1993.
* 1990 estaba en \PremCopied pero el Gini no replica
* Los otros 4 años (96,99,02 & 04) que también estaban \PremCopied y se guardaron en vintage control.
* KGZ 1998 -> La base guardada en \PremCopied dice que el welfare es **Ingreso**, pero povcalnet dice que es **Consumo**. El Gini replica exacto.
  + LAO 1992 -> La base guardada en \PremCopied no replica el Gini.
  + LKA 1990 & 1995 -> Las bases guardada en \PremCopied no replica el Gini.
* RENOMBRAR DOFILES -> A (ALB) - L (LAO)
* LVA 1998 -> el dta de 1998 en \PremCopied tenía year=1997, pero los estimados corresponden al 98. Cambié el año al guardar en GMD
* MEX:
  + 1984 es definitivamente Ingreso. Hay que cambiar en el Master, povcalnet y en todas partes
  + La variable hhid en 84 y 89 es la raw en str, pero habría que transformarla
  + R// se guardó la última versión de MEX 84 bajada del NSO como v02.
* MKD
  + 2009 no aparece como diferencia entre povcalnet & dwb, pero no hay folder para este año en Vintage.
* MLI -> 1994 Cambié el coverage de Rural a Nacional en el master surveyinfo sheet. En Povcalnet web también aparece mal ([link](http://iresearch.worldbank.org/PovcalNet/Docs/CountryDocs/MLI.htm#3)). Pero povcalnet command y documentos oficiales indican que esta encuesta es nacional (9700 hhs). El dta en el folder de Prem también tiene valores 1 y 0 en variable urban. <http://www.instat-mali.org/nada/index.php/catalog/9/study-description>
* MYS
  + Los años y siglas de la encuesta en Vintage control son inconsistentes con lo que hay en povcalnet (HIS 2009, 12, 14, 16 vs. HIBAS 2008, 11, 13, 15).
* NGA
  + 1992 -> original dta had year=1992.5, but povcalnet has year & datayear as 1992 only.
  + 1996-> all records indicate that is National, and Gini replicates well, but 100% of obs are rural.
* SUR 1999
  + El dta en \PremCopied no replica el Gini. Encontré otro dta que replica el Gini en otro folder (**P:\Documentation\Surveys\SUR**\SUR\_N1999Y\Data)
  + Todas las referencias indican que la encuesta es Nacional, pero el 100% de las observaciones son urbanas.
* Hay varias encuestas cuyo peso (weight) es igual al hhsize: Ej: SUR 1999, SYC 1999, SYR 2004, UKR 92, 93 & 96.
* SCY 1999: Registros indican que la cobertura es nacional, pero 100% de las observaciones en urban son 0.
* VEN -> ninguna de las bases en /PremCopied replica. La data está claramente a nivel individual, pero la variable hhid está mal (aparece un hh id por observación), y hay diferentes pesos en personas de un mismo hogar
  + 5/24: Las encontré todas en P:\Kihoon\Handover\Data\LAC\DS\SEDLAC-Total

Los ginis replican PERO **SIN RESTRINGIR POR COHH**. Las bases guardadas en el folder de prem son cohh==1, por eso el Gini en povcalnet no replica con esas.

* Alguna razón por la cual y2000 no está en povcalnet? El folder de Kihoon tiene esa base.
* ZAF 2008 -> Todos los records indican que la encuesta es nacional, pero 100% de las obs son rurales.