This dataset contains the raw data of questionnaire named “Original data” under the file “Questionnaire\_data\_\_butterfly\_wing\_color\_brightness\_and\_contrast.xlsx”, that is the responses of local Dai people to the twenty questions (Questions were indexed as the abbreviation of the categories from BK1 to AK5. See the second column on “Proportion”). This data was used to do Kaiser-Meyer-Olkin (KMO) and Bartlett’s Test of Sphericity in SPSS 17.0, and Confirmatory factory analysis (CFA) in SPSS Amos Graphics (see supplementary materials in Yang et al, “Biased local ecological knowledge of butterfly diversity provided by the ethnic minority in Southwest China’s biodiversity hotspot”).

The file named “butterfly-images.rar” is the digital images of butterfly which were used to measure the wingspan, wing color brightness and contrast. These data were devoted to analyses the divergency between the butterflies recognised or not by local Dai people. The data stored in the “Questionnaire\_data\_\_butterfly\_wing\_color\_brightness\_and\_contrast.xlsx” file and sheets named “Wing color brightness data”, “Wing color contrast data” and “Wing span data” respectively.

The questionnaire was divided into five categories related to the conservation awareness and perception of the butterflies: basic knowledge about butterflies (BK), perspectives on changes in local butterfly diversity (BD), knowledge about human and environmental impacts on butterfly diversity (HI), knowledge about cultural significance of butterflies (CS), awareness and knowledge about butterfly conservation (AK). CSV Files named “BK1 to HI2” are the response of local Dai people to the 20 questions. These files were used to do cumulative link mixed models (CLMM) test whether gender and age affect response patterns of the individual questions. Among the 20 csv files, “Sex.1M.2F” refer to gender, 1 is Male and 2 is Female; “Age1J.2Y.3S” refer to age categories. The ages of the respondents were grouped into youth (15-22 years old), younger-adult (23-45 years old) and older-adult (45-65 years old) and signed as 1 to 3 in the dataset. Respondents from five villages also signed as 1 to 5. Detail in supplementary materials in Yang et al, “Biased local ecological knowledge of butterfly diversity provided by the ethnic minority in Southwest China’s biodiversity hotspot”).

The data related to demographic difference and proportion of local respondents, the results of CLMM and Chi-square tests were also stored in the “Questionnaire\_data\_\_butterfly\_wing\_color\_brightness\_and\_contrast\_etc.xlsx”.