*** PROGRAM TO ESTIMATE DEMOGRAPHIC DETERMINANTS OF CHILDCARE PREFERENCES ***

1. Clean GSS-Family data to extract or create variables relevant for analysis.
   1.1. Select cases where Respondent:
      1.1.1. has children (TOTCHDC);
      1.1.2. at least one of those children lives with Respondent at least part-time (HHDSTA*);
      1.1.3. at least one of those children is 12 years of age or less (ACHD_*C).
   1.2. Select demographic variables – Respondent's age (AGEC), gender (SEX), marital status (MARSTAT), family income (FAMINCG2), education (EHG3_01B), place of birth (BRTHCAN), visible minority status (VISMIN), and participation in religious activities (REE_02) – and the child care variables (CC_10_*, CC_20_*, CC_30_*, CC_40_*, CP_10_*, CP_20_*).
   1.3. Create the dependent variable (cc_pref) in a new data frame (final_data), such that the categories of that variable are the same as those in the existing child care preference variables (CP_20_*), but where, if the Respondent *doesn't* want to change their current arrangements (CP_10_*), the current child care arrangement variables (CC_10_*, CC_20_*, CC_30_*, CC_40_*) are used to assign equivalent categories. Given that can be up to 7 such preferences expressed, take the preference for the *last* child listed (presumed to be the youngest) as the single preference for the Respondent.
   1.4. Recode demographic variables into the new data frame such that:
      1.4.1. binary variables (gender, place of birth, visible minority) are Boolean (Female, Canada, Yes as TRUE);
      1.4.2. marital status is collapsed and reordered to Single; Married OR Common-law; Widowed, Divorced or Separated;
      1.4.3. participation in religious activities is put in reverse order, from Not at all to At least once a week;
      1.4.4. Valid skip, don't know, Refusal, and Not stated values become missing values (NA).
   1.5. Drop any cases where the dependent variable is missing.

2. Generate tables of summary statistics for all variables: (t)
   2.1. five-number summary with mean and standard deviation for numeric data;
   2.2. frequencies and proportions for all categories (including missing, if any) otherwise.

3. Run a multinomial logistic regression of childcare preferences on the included demographic variables, with Parent care as the baseline, and a type II analysis of deviance of the results to identify the "significant" variables.

4. Generate summary tables of the regression coefficients and the analysis of deviance. (t)