Practical session

GCC-STAT Regional Workshop

Muscat, Oman, 22-24 September 2019

Deliveries (1)

- Database structure
  - Variable name
  - Description (content) of the variable
  - Format
    - Number
    - Character (varchar2)
  - Length
  - Classification used

Deliveries (2)

- Classifications (14)
  - Code
  - Description

Deliveries (3)

- Working database
  - 27 variables
  - 51,680 records
  - Several method of statistical disclosure control used
    - Re-coding
    - Aggregation
    - Supplementary key method
    - Adding noise to data

Task 1

- Calculate correction rate and imputation rate for the variable DW_ID (dwelling number)
- Correction rate – 4.2%
  - DW_ID_CRP = not missing AND (DW_ID_CRP<> DW_ID)
- Imputation rate – 2.1%
  - DW_ID_CRP = missing

Task 2

- Prepare rules for automated correction of marital status (MAR) where MAR = 9 and check consistency between age and marital status

- Use age (persons below 15 years)
  - International standard threshold
    - IF AGE= 0-14 AND MAR =9 THEN MAR = 1 (260 out of 1,772)
- Use of data from variable HH_STAT
  - IF HH_STAT = 01, 02 AND MAR = 9 THEN MAR = 2 (278 out of 1,512)

Task 2 (continued)

- Statistical presumptions
  - Children living with parent(s) are single
    - IF HH_STAT = 07-10 AND MAR = 9 THEN MAR = 1 (139 out of 1,234)
  - Persons younger than 30 years are single
    - IF AGE = 15-29 AND MAR = 9 THEN MAR = 1 (201 out of 1095)

Task 2 (continued)

- Imputations (for rest 894 records)
  - Stratum: CIT
  - Matching variables: SEX, AGE
- Consistency check
  - 2 records with influential error
    - AGE = 0-14 AND MAR = 2

Task 3

- Detect outliers and influential errors for labour force status data (ACT) before editing

Task 3

AGE	ACT										
	EMP	EMP	EMP	UN_E	UN_E	CH	PUP	STU	RET	OTH	OTH
0						707					
1-4						2515					
5-9						2596					
10-14	1		1			2093	1	1			1
15-19	88	5	1	42	5		2103	179		51	259
20-24	1070	53	8	106	79		146	1255		88	312
25-29	2893	127	28	147	173			207		121	485
30-34	3275	180	29	56	239			25		102	475
35-39	2980	196	28	39	201			4		68	422
40-44	2830	181	37	29	189			4	21	50	410
45-49	2634	191	82	18	225			2	30	39	370
50-54	2422	199	134	13	223				77	30	362
55-59	1551	125	114	7	256				433	33	292
60-64	427	61	43	1	121				1154	13	224
65-69	76	14	5						1416	2	158
70-74	11	2	4						1054	1	84
75-79									824	2	47
80-84			3						508		16
85+	1							1	416		17

Task 4

- Detect households with farmers (ACT = 03) and find out how many household members have no data on labour force status
  - 459 households
  - 394 persons

Task 5

- What would be strategy to solve missing data for the variable industry (IND)
  - Automated correction for farmers
    - IF ACT = 03 AND IND = null THEN IND = A (25 out of 4,947)
  - Imputation for employed working in Slovenia (2,183)
    - IF ACT = 01, 02 AND IND = null AND POW = 1 THEN
      - Stratum: MUN
      - Matching variables: SEX, AGE
      - Threshold: 20
  - No imputation for employed working abroad (2,737)

Task 6

- Define stratum and matching variables for imputation educational attainment (EDU) data and check the minimum threshold (20 records)
  - Stratum: CIT
  - Matching variables: SEX, AGE

Task 6 (continued)

- Imputations for which sub-population will not be executed
  - Older population
  - EU and other country citizens
- Possible improvement of imputation rules
  - Aggregation of age groups / citizenship