# CHAPTER IV

# VALUATING THE EFFICIENCY OF COMPLEX SAMPLE DESIGNS FOR THE MEDIA INDEX SURVEY

## Characteristics of Households and Respondents

The Media Index Survey (MIS) elicited a host of information on the socio-
ɔgraphic background and media consumption of the sample population. This
'mation is needed for a detailed study of the differentials in media penetration.
ıographic and background variables such as age, gender, educational level and
icity are important explanatory variables in any socio-economic and market
eys.

For an initial assessment of the representativeness of the MIS sample, the
ple distribution on a number of background variables are compared to the 1998
nates of the total population. Table 4.1 shows that the ethnic distribution of the
ı sample matches rather closely that of the total population in Peninsular
aysia. However, the sample distribution by age group was skewed towards older
ulation as compared to the 1998 population estimates (DOS, 1998). Those aged
/ears and above were over-represented by about 5 percentage points in relation to
ɔorresponding population. Further investigation shows that over-representation
ılder persons in the sample occurs for both males and females.

The MIS was designed in such a way that each quarter is an independent
ıple and identical between the fieldwork quarters. As such, another way of
luating the reliability and consistency of the sample is to compare its distribution
:r the four fieldwork quarters. Table 4.2 shows that the sample distributions for
h fieldwork quarter are fairly consistent with respect to ethnic group and

ehold income. However, some fluctuations can be observed in the sample ibution by type of dwelling units, especially for the second quarter of 1998 as pared to the other 3 quarters. The fluctuation may be due to some biases in the ction of PSUs for that quarter.

Overall, nearly 40% of the sample households reported a monthly income of 1000 and below. The proportion of households with monthly income below the 1000 level is significantly higher in the rural areas as compared to the urban and ropolitan strata, especially the latter. At the other end of the income spectrum, about 4% of the households reported a monthly income of RM5000 and above, 75% of these 'affluent' households were from MT stratum (see Table 4.3). me data are generally known to be less accurate due to under-reporting. About f every 100 respondents in each quarter refused to disclose household income rmation, and slightly more than half of these were Chinese respondents, while Malays made up some 36 percent.

**Table 4.1: Percent Distribution of Media Index Sample and Population Estimates 1998**

| | Media Index Sample | | Estimates 1998 | |
|---|---|---|---|---|
| **Race** | | | | |
| Malays | 59.5 | | 60.0 | |
| Chinese | 29.0 | | 28.0 | |
| Indians/Others | 11.5 | | 12.0 | |
| Total | 100 | | 100 | |
| **Age Group** | | | | |
| 15-29 years old | 30.9 | | 40.8 | |
| 30-49 years old | 44.2 | | 38.7 | |
| 50+ years old | 24.9 | | 20.5 | |
| **Age Group** | Male | Female | Male | Female |
| 15-29 years old | 31.6 | 30.3 | 41.5 | 40.0 |
| 30-49 years old | 43.5 | 44.7 | 38.4 | 39.0 |
| 50+ years old | 24.9 | 25.0 | 20.1 | 21.0 |

**Table 4.2: Percent Distribution of Sample by Fieldwork Quarter**

| | Quarter | | | | |
|---|---|---|---|---|---|
| | Q3/97 | Q4/97 | Q1/98 | Q2/98 | Total |
| e | | | | | |
| ays | 61.4 | 55.3 | 61.0 | 59.9 | 59.5 |
| iese | 27.6 | 32.3 | 26.4 | 30.0 | 29.0 |
| ans/Others | 11.0 | 12.4 | 12.6 | 10.1 | 11.5 |
| il | 100 | 100 | 100 | 100 | 100 |
| **isehold Income** | | | | | |
| applicable* | 2.0 | 4.2 | 3.1 | 4.7 | 3.5 |
| o RM300 | 4.5 | 3.4 | 3.8 | 5.2 | 4.2 |
| 301-RM500 | 8.5 | 8.3 | 8.0 | 9.0 | 8.5 |
| 501-RM750 | 11.1 | 11.8 | 11.9 | 9.7 | 11.1 |
| 751-RM1000 | 13.1 | 13.1 | 12.7 | 11.8 | 12.7 |
| 1001-RM1500 | 16.2 | 15.7 | 15.3 | 15.4 | 15.7 |
| 1501-RM2000 | 11.2 | 12.6 | 11.4 | 12.3 | 11.9 |
| 2001-RM2500 | 8.0 | 6.9 | 7.2 | 6.7 | 7.2 |
| 2501-RM3000 | 6.2 | 6.6 | 7.2 | 5.9 | 6.5 |
| 3001-RM4000 | 6.7 | 4.4 | 6.4 | 7.0 | 6.1 |
| 4001-RM5000 | 3.3 | 3.7 | 3.4 | 3.3 | 3.4 |
| 5001-RM8000 | 2.9 | 2.3 | 3.0 | 2.2 | 2.6 |
| ir RM8000 | 1.3 | 1.4 | 2.0 | 1.4 | 1.5 |
| stated | 4.8 | 5.6 | 4.6 | 5.3 | 5.1 |
| al | 100 | 100 | 100 | 100 | 100 |
| **ie Dwelling Unit** | | | | | |
| igalow/Detached/Compound | 2.2 | 2.3 | 3.5 | 4.2 | 3.1 |
| ni-detached | 3.9 | 4.2 | 5.1 | 3.8 | 4.3 |
| race/Link-2 or more storey | 13.0 | 11.1 | 12.1 | 13.1 | 12.3 |
| race/Link -single storey | 26.1 | 20.6 | 25.5 | 13.9 | 21.5 |
| ip-house | 1.2 | 0.7 | 0.4 | 0.9 | 0.8 |
| v cost house | 6.1 | 9.6 | 5.8 | 7.7 | 7.3 |
| v cost flat | 2.0 | 3.6 | 2.3 | 2.7 | 2.6 |
| dium cost flat | 4.1 | 5.1 | 5.1 | 5.5 | 5.0 |
| art. Above shops | 0.1 | 0.5 | 0.0 | 0.6 | 0.3 |
| xury apart./flat/condo | 0.1 | 0.5 | 0.3 | 2.7 | 0.9 |
| ap/Zink-village/Squatters | 40.8 | 41.4 | 38.4 | 44.7 | 41.2 |
| iers | 0.5 | 0.3 | 1.6 | 0.3 | 0.7 |
| tal | 100 | 100 | 100 | 100 | 100 |
| **mber of households** | **2199** | **2084** | **2237** | **2218** | **8738** |

his group of respondents was supported by other family members that were not taying together with the respondents.

**Table 4.3: Percent Distribution of Household Income by Stratum**

| | MT | UT | RT | Total |
|---|---|---|---|---|
| pplicable | 1.8 | 4.1 | 4.7 | 3.5 |
| RM300 | 0.6 | 2.9 | 9.2 | 4.2 |
| 01-RM500 | 1.6 | 7.1 | 16.6 | 8.5 |
| 01-RM750 | 4.9 | 10.9 | 17.7 | 11.1 |
| 51-RM1000 | 8.2 | 13.4 | 16.4 | 12.7 |
| 001-RM1500 | 14.6 | 18.2 | 14.3 | 15.7 |
| 501-RM2000 | 14.6 | 13.3 | 7.9 | 11.9 |
| 001-RM2500 | 10.1 | 7.4 | 4.2 | 7.2 |
| 501-RM3000 | 10.3 | 6.4 | 2.7 | 6.5 |
| 001-RM4000 | 10.1 | 5.9 | 2.4 | 6.1 |
| 001-RM5000 | 6.7 | 2.9 | 0.7 | 3.4 |
| 001-RM8000 | 5.4 | 1.9 | 0.4 | 2.6 |
| RM8000 | 3.8 | 0.4 | 0.3 | 1.5 |
| tated | 7.3 | 5.2 | 2.6 | 5.1 |
| | 100 | 100 | 100 | 100 |
| **ber of Respondents** | **2985** | **2798** | **2955** | **8738** |

Table 4.4 shows the distribution of respondents by gender and age group. overall sex ratio from MIS is 81 males per 100 females. Compared to the total lation with approximately equal number of males and females (DOS, 1998), figure shows an under-representation of the males in the sample. This may be uted to the lower response rate of males as compared to the females, as the er tends to be less likely to be at home as compared to the latter. Data show only 2652 male respondents were interviewed on the first visit compared to 5 female respondents for the same attempt.

The age composition of the sample is quite similar for both sexes within each um. The younger age structure of the MT stratum as compared to the RT um could be attributed to rural-urban migration, which is the selective of the nger population.

**Table 4.4: Percent Distribution of Respondents by Stratum and Selected Variables**

| | MT | UT | RT | Total |
|---|---|---|---|---|
| group | | | | |
| ) | 33.9 | 32.5 | 28.2 | 31.6 |
| ) | 47.1 | 43.1 | 39.7 | 43.5 |
| | 19.0 | 24.4 | 32.2 | 24.9 |
| | 100 | 100 | 100 | 100 |
| **l Male** | **1449** | **1185** | **1278** | **3912** |
| **ale** | | | | |
| group | | | | |
| 9 | 33.4 | 29.3 | 28.3 | 30.3 |
| 9 | 46.2 | 47.2 | 41.1 | 44.7 |
| | 20.4 | 23.4 | 30.6 | 25.0 |
| l | 100 | 100 | 100 | 100 |
| **l Female** | **1536** | **1613** | **1677** | **4826** |
| **)verall Sex Ratio** | **0.94** | **0.73** | **0.76** | **0.81** |
| **Educational Level** | | | | |
| ormal education | 4.5 | 9.0 | 15.3 | 9.6 |
| ıary | 19.8 | 26.8 | 35.7 | 27.4 |
| er Secondary | 22.1 | 20.5 | 20.5 | 21.0 |
| dle Secondary | 31.9 | 29.8 | 22.0 | 27.9 |
| er Secondary/HSC | 7.1 | 5.3 | 3.4 | 5.3 |
| ege | 8.3 | 5.0 | 1.8 | 5.0 |
| versity | 5.9 | 3.6 | 1.3 | 3.6 |
| ısed | 0.3 | 0.0 | 0.1 | 0.1 |
| ıl | 100 | 100 | 100 | 100 |
| **ıber of Respondents** | **2985** | **2798** | **2955** | **8738** |

The level of educational attainment in a society is an important indicator of ial development. Moreover, education is considered to be a major factor erlying social status that ultimately impacts on the attitude towards and usage of ious consumer products. Overall, about 76% of the respondents had at least ldle secondary schooling or below. Only about 10% of the respondents had

attended school (see Table 4.4). The majority of those with no schooling ) were senior citizens aged 50 years old and above and mostly residing in the areas. Consistent with findings from population censuses, the educational level g urban respondents is significantly higher than that of their rural counterparts.

Table 4.5 shows that overall, about 66% of the respondents were working , and about 28% of non-working adults were housewives. Blue-collar workers sent the largest group of working adults.

In summary, the MIS sample is approximately reflective of the total lation of Peninsular Malaysia, and the sub-samples over the four quarters are quite consistent with one another. The sample covered respondents in all the raphical areas with various socio-demographic characteristics that are rather sentative of the Malaysian society.

**Table 4.5: Percentage Distribution of Respondents by Occupation**

| Occupation | Percentage |
|---|---|
| Professional & Executive | 5.2 |
| Businessman - Big & Small | 4.0 |
| Other White-Collar / Skilled Worker | 14.0 |
| Student | 10.3 |
| Blue-Collar (Semi Skilled/Skilled) | 15.5 |
| Labourer | 14.1 |
| Farmer & Fisherman | 2.7 |
| Unemployed | 1.7 |
| Female Housewife | 28.1 |
| Retired | 4.2 |
| Not Stated | 0.2 |
| **Total Sample** | **8738** |

## 4.2 Exploratory Data Analysis

The Media Index Survey (MIS) contained numerous variables ranging from usage and attitude of individual to household products or services. Besides, the survey also elicited information on the ownership of a large number of consumer durable items. In view of the numerous variables available in the data set, it is impractical and unnecessary to compute and present the design effect for every single variable in this report. Therefore, a set of variables will be selected for the computation of sampling errors and design effects. A subset of above variables will then be identified for a more detailed subclass analysis of design effects as well as the coefficients of intra-class correlation.

The characteristics and sample distribution of the subset of variables were examined below to provide some insights on each of them. The exploratory analysis is aimed at providing some insights for a better understanding of the design effects and the coefficients of intra-class correlation in the subsequent section.

Figures 4.1 to 4.3 illustrate the ownership of 3 household durable goods, and 2 household daily products bought in the past 1 month for MT, UT and RT respectively. The household durable goods are telephone (DURA12), refrigerator (DURA29) and air conditioner (DURA35). Telephone penetration rate of the households in Peninsular Malaysia is 72.5%. But, the penetration rate is not evenly distributed across geographical domains. The penetration rate is 84% in MT, 77% in UT and 57% in RT. Besides, the penetration rate also varies substantially across ethnic groups between and within strata. For example, only slightly more than half of the Malays and Indians/Others households in RT stratum have telephone as compared to 80% for Chinese households in the same stratum.

Data show that 88% of the households own at least one refrigerator with little variation across ethnic groups. As expected, the penetration rate of refrigerator is slightly lower in RT compared to the other two urban strata. The ownership of air conditioner is still very low among households in Peninsular Malaysia. Overall, only 14% of households have air conditioners. Of the 1259 households with air conditioners, 74 percent are Chinese households, and 773 units are in the MT stratum.

The 2 daily consumer products bought in the past 1 month to be analyzed in this report are concentrated detergents (P1MBUY15) and baby diapers (P1MBUY18). Concentrated detergent is a common household's cleaning product used by all segments of the population. Figures 4.1 to 4.3 show little variation in the purchase of this product across ethnic groups and strata. However, only about 40% of the households were using concentrated detergent as there are other cleaning products available in the market. The purchase of baby diapers in the past 1 month is closely related to recent fertility rate and household income as baby diapers are selling at a premium price in the market. Overall, 12% of the households reported having bought baby diapers in the past 1 month and the rate is about the same for both urban and rural areas. Chinese households with a total fertility rate of only 2511 per 1000 women as reported in Vital Statistics Malaysia 1998 (DOS, 1998 p52) are less likely to purchase the diapers as compared to the other ethnic groups with higher fertility. Reflecting their higher fertility rate, Malay households in all strata are most likely to buy baby diapers in the past 1 month.

Figure 4.1: Products Penetration Rate by Ethnic Group in Metropolitan Towns (MT)

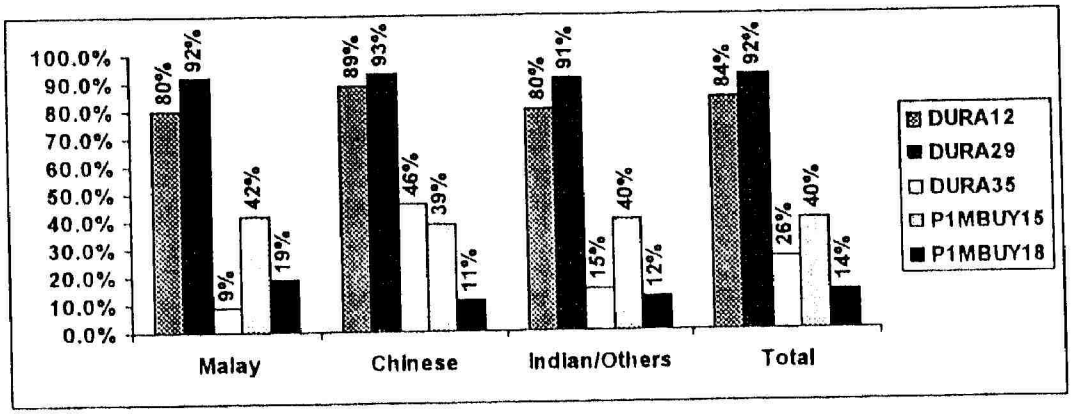



Figure 4.2: Products Penetration Rate by Ethnic Group in Urban Towns (UT)

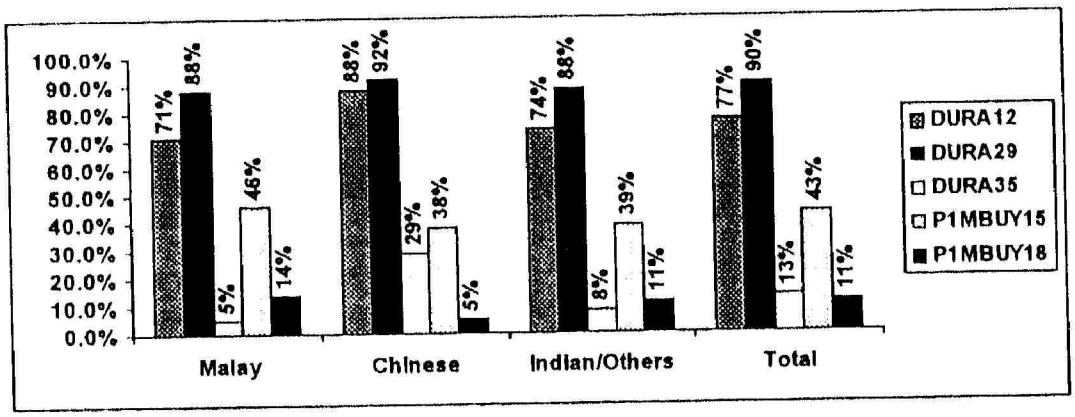
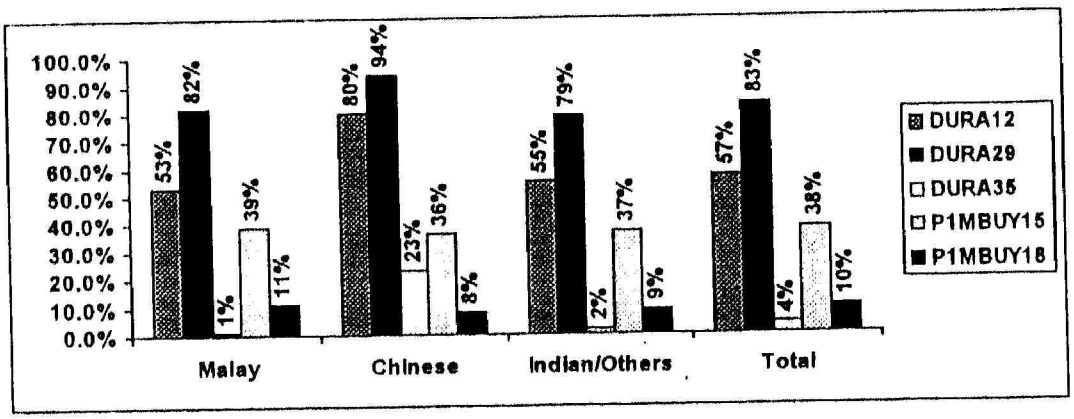



Figure 4.3: Products Penetration Rate by Ethnic Group in Rural Areas (RT)



Of the individual variables, exploratory analysis is done on (i) the percentage of individual who claimed to have participated in sport during leisure time (ACTVT1), (ii) the ownership of ATM card (BANK13), (iii) literacy rate in Bahasa Malaysia (LITER3) and (iv) the percentage who had read any newspapers the day prior to the survey (NPYTDA).

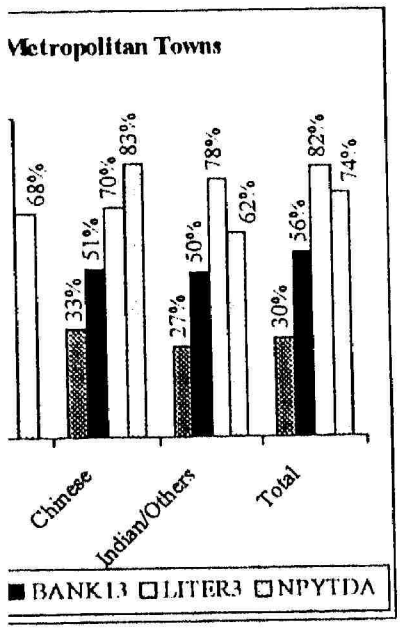
Figure 4.4 shows the percentage of respondents who had a particular attribute according to the four selected variables mentioned above, cross-classified by ethnic group and stratum. Participating in sports activity is one of the leisure activities among Malaysians, besides reading newspapers and magazines, visiting friends, watching television and listening to the radio. In the MIS sample, only one in four respondents reported having participated in sports. Metropolitan respondents are most likely to have participated in sports while the rural respondents are the least likely. However, participation rate in sports does not seem to vary by ethnicity.

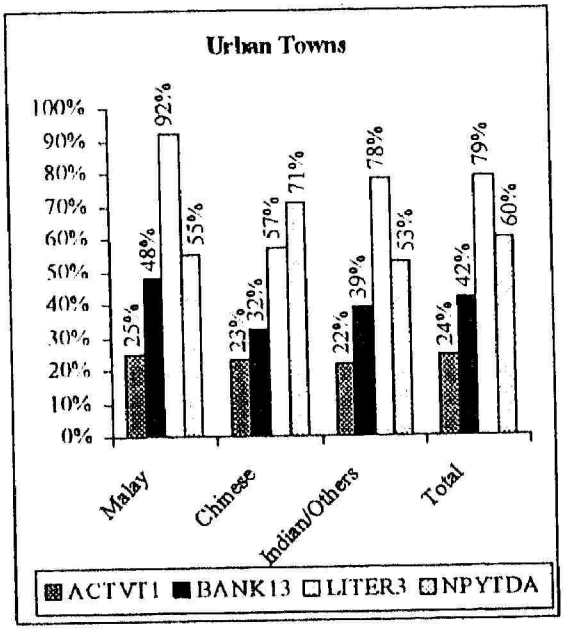
About 43% of adults in the sample claimed to own at least one ATM card. The ownership rate is highest in MT stratum (56%), followed by UT stratum (42%) and RT stratum (29%). Ownership of ATM cards by ethnic groups varies by stratum, but the overall ethnic differentials in ownership of ATM card varies within a narrow range of between 41% and 44%.

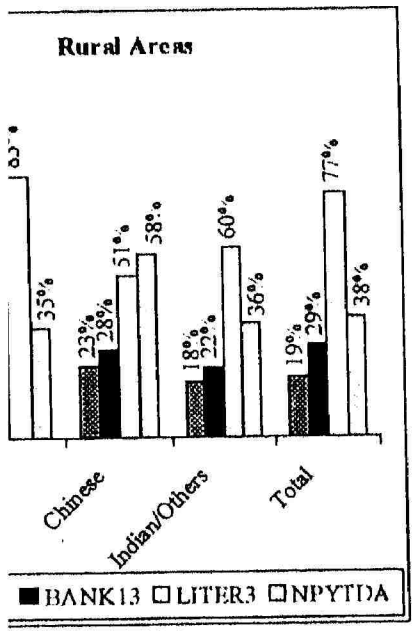
Overall, 89% of Malay respondents are literate in Bahasa Malaysia followed by Indians/Others (74%) and Chinese (62%), with the highest rate in MT and the lowest rate in RT. Chinese and Indians/Others have a slightly lower literacy rate in Bahasa Malaysia but their literacy rate in English may be higher than that of the Malays, and certainly in their own language. Fully 88% of Chinese and 98% of Indians/Others in the sample claimed to be able to speak Bahasa Malaysia.

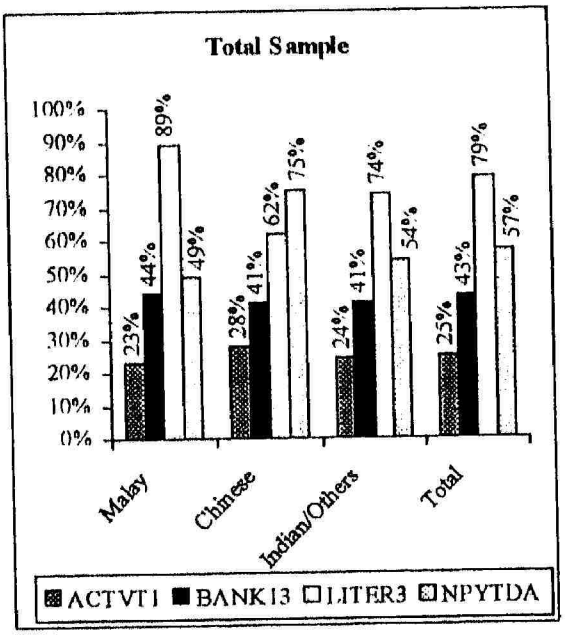
Even though only 10% of the respondents claimed that they did not attend ny formal schooling, newspaper readership is not very high among the respondents. Only about 57% of the adults interviewed claimed that they had read any newspapers the day prior to the survey. Again, newspaper readership is much higher n MT, especially compared to RT stratum. Among Malays and Indians/Others, newspaper readership in MT (68% and 62%) is almost twice the rate in RT stratum (38% and 36%). About 83% of Chinese in MT read any newspapers the day prior to the survey, and there is little difference between urban and rural areas. Overall, some 72% of the respondents reported that they did read any newspapers in the past 7 days prior to the survey.

**Figure 4.4: Percentage of Respondents of Selected Individual Variables by Ethnic Group and Stratum**



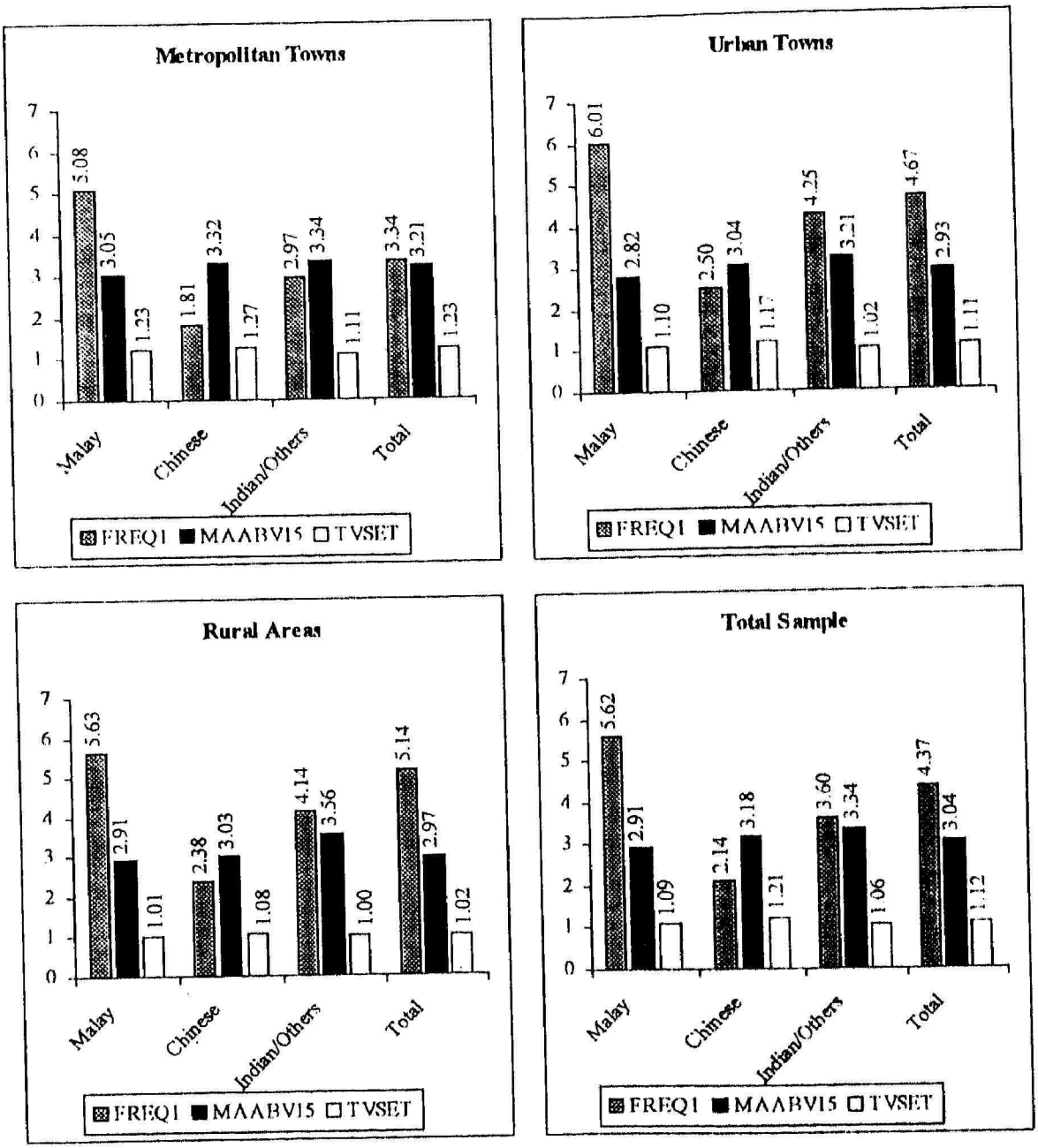



Exploratory analysis is also done on the mean values of the following three variables: (i) number of days in a week that respondents turned on TV1 (FREQ1), (ii) number of members aged 15+ years old in the household (MAABV15), and (iii) number of television sets in the household (TVSET). Figure 4.5 illustrates the mean values of these variables by stratum and ethnic group.

The average number of days in a week the respondents turned on TV1 programs (FREQ1) ranges from 3.34 days in MT stratum to 4.67 days in UT stratum and 5.14 days in RT stratum. The higher mean value in the RT stratum may be explained by the fact that the Malays who make up the largest group in rural areas are more likely to turn on TV1 programs, with a mean of 5.62 days per week. The Indians/Others, especially those staying in UT and RT strata, also turned on TV1 frequently, with an average of 4.25 days and 4.14 days per week respectively. Overall, the number of days the Chinese turned on TV1 programs averaged only 2.14 days a week, with an average of only 1.81 days among those from the MT. The Indians/Others in MT stratum are also less likely than the Malays to turn on TV1 programs (average of 2.97 days), as compared to their counterparts in UT and RT strata. These findings are not surprising, as there are other TV channels, which are more attractive for the Chinese and Indians/Others. Ability to subscribe to cable and satellite television is another factor that reduces the exposure to TV1, especially for those who stay in MT stratum. In addition, the residents of southern region of Peninsular Malaysia can also view television programs from Singapore.

**Figure 4.5: Mean Values of Selected Individual Variables by Ethnic Group and Stratum**



Data show that, on average, there are 3.04 adults aged 15 years and above per household (MAABV15) in the MIS sample. The average number of adults is larger for households in MT stratum (3.21), and slightly below 3 for households in UT and RT strata. This may be explained by rural-urban migration among the population, involving mainly the young people.

In terms of ethnicity, Indians/Others have the largest average number of adults in the households in all strata compared to Malays and Chinese. Greater difference is observed in the RT stratum where the average number of adults in Indians/Others households is 3.56, compared to 3.03 for the Chinese and 2.91 for the Malays. The Malay households have smaller number of adults but they have more members who are below 15 years of age. The average number of members aged below 15 years old for Malay households is 1.81 (MABLW15), as compared to 1.49 for the Indians/Others, and 1.13 for the Chinese in the overall sample.

Television ownership is almost universal in Peninsular Malaysia, with an average of 1.12 TV sets per household. About 13% of the households in MIS sample have multiple TV sets. Table 4.6 shows that the number of TV sets in a home is closely correlated with household income.

In summary, the above 12 variables were selected to represent a large number of variables obtained in the MIS measuring different aspects of purchasing and usage behavior. These selected variables are related to socio-demographic variables in different ways, and they provide a wide range of scenario for a comprehensive study of the design effects and rate of homogeneity. Some of the selected variables are normally distributed (DURA29, MAABV15 and P1MBUY18), some are likely to be urban skewed (DURA35, BANK13 and NPYTDA) and some may differ widely by ethnic groups (FREQ1 and LITER3).

Some of the variables are also used as proxies of the socio-economic status of the households (TVSET, DURA35).

**Table 4.6: Number of TV Sets in The Household by Household Income**

| | Number of TV Sets | | | | | | |
|---|---|---|---|---|---|---|---|
| **Household Income** | **0** | **1** | **2** | **3** | **4** | **5** | **Total** |
| Not applicable | 40 | 251 | 15 | 1 | | | 307 |
| Up to $300 | 79 | 282 | 9 | | | | 370 |
| $301-$500 | 47 | 671 | 21 | | | | 739 |
| $501-$750 | 34 | 894 | 42 | 3 | | | 973 |
| $751-$1000 | 30 | 993 | 78 | 6 | | | 1107 |
| $1001-1500 | 22 | 1200 | 136 | 8 | 2 | | 1368 |
| $1501-$2000 | 13 | 904 | 111 | 8 | 2 | | 1038 |
| $2001-$2500 | 3 | 512 | 97 | 15 | 3 | | 630 |
| $2501-$3000 | 5 | 466 | 81 | 11 | 2 | 1 | 566 |
| $3001-$4000 | 4 | 385 | 132 | 12 | 3 | 1 | 537 |
| $4001-$5000 | 5 | 197 | 82 | 16 | 1 | | 301 |
| $5001-8000 | 2 | 135 | 68 | 18 | 2 | 1 | 226 |
| Over $8000 | 1 | 53 | 54 | 18 | 7 | 1 | 134 |
| Not stated | 6 | 350 | 64 | 18 | 4 | | 442 |
| TOTAL | 291 | 7293 | 990 | 134 | 26 | 4 | 8738 |

## 4.3 The Patterns of Design Effects

Sampling errors and design effects (*deft*) covering a large number of variables concerning media consumption, banking services, household size, household durable goods, and daily consumer products that the households bought in the past one month are computed for total sample and by stratum.

The following Table 4.7 illustrates an example of calculating the standard errors of variable *y* using Taylor Expansion method.

**Table 4.7: Example of Standard Error Computation**

| A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|
| Stratum | psu | $x_i$ | $a_h$ | $a_h/(a_h-1)$ | $y_i$ | $z_{hi}$ and $z_h$ | $z_{hi}^2$ and $\text{sum}(z_{hi}^2)$ | $\text{sum}(z_{hi}^2) - z_h^2/a_h$ | E * I |
| 1 | 1 | 8 | | | 8 | 0.82 | 0.67 | | |
| 1 | 2 | 14 | | | 14 | 1.43 | 2.06 | | |
| 1 | 3 | 18 | | | 17 | 0.84 | 0.71 | | |
| Total stratum 1 | | 40 | 3 | 1.50 | 39 | 3.09 | 3.44 | 0.24 | 0.36 |
| 2 | 1 | 24 | | | 22 | 0.46 | 0.21 | | |
| 2 | 2 | 37 | | | 34 | 0.79 | 0.62 | | |
| Total stratum 2 | | 61 | 2 | 2.00 | 56 | 1.25 | 0.83 | 0.05 | 0.11 |
| 3 | 1 | 23 | | | 17 | -3.64 | 13.28 | | |
| 3 | 2 | 18 | | | 14 | -2.16 | 4.65 | | |
| 3 | 3 | 5 | | | 5 | 0.51 | 0.26 | | |
| 3 | 4 | 19 | | | 18 | 0.95 | 0.89 | | |
| Total stratum 3 | | 65 | 4 | 1.33 | 54 | -4.34 | 19.09 | 14.37 | 19.17 |
| TOTAL | | 166 | | | 149 | | | | 19.64 |
| $r = y / x$ | | 0.90 | | | | | | var(r) = | 0.0007 |
| | | | | | | | | se(r) = | 0.0267 |

Assume that there are 3 strata with stratum 1 has 3 PSUs selected, stratum 2 has 2 PSUs selected and stratum 3 has 4 PSUs selected. Let variable *y* measure the ownership of telephone in the household. Then, the total number of interviews and total number of households owned telephone are obtained for every PSUs as in column C and column F, respectively. The total number of selected PSUs by stratum ($a_h$) is shows in column D and column E gives the value of $a_h/(a_h-1)$ by stratum.

The ratio estimate, r = 149/166 = 0.90 gives the ownership of telephone of 90% in this example. Subsequently, $z_{hi} = y_{hi} - r*x_{hi}$ is computed for every PSUs and $z_h = y_h - r*x_h$ for every strata (column G). Then, $z_{hi}^2$ is computed for every PSUs with the total value, sum($z_{hi}^2$) by stratum as shown in column H. The value of (sum($z_{hi}^2$) - $z_h^2 / a_h$) is then computed for every strata (column I). The column J shows the value of multiplying column E and column I by stratum. Lastly, the variance for r is obtained with var(r) = (0.36 + 0.11 + 19.17) / $166^2$ = 0.0007 and the se(r) = 0.0267.

The overall sampling fraction is omitted in above calculation because it is very small and also for the ease of calculation. The above process is repeated to compute the standard error for every variable in this study. An example of the actual calculation worksheet for variable DURA12 is attached in Appendix II.

A total of 60 variables for household and individual characteristics are presented for total sample in this report. It is necessary and useful to present these variables so as to be able to identify patterns of variation in the results. Besides, the variables are grouped into 5 categories namely 'Variable Category' with each consisting of variables measuring approximately similar dimensions or attributes. Table 4.8 illustrates the grouping and list of variables in each category. The variable names in bold indicate selected variables for which exploratory analysis was done in Section 4.2. The analysis of design effects (*deft*) will focus on these variables within the geographical domains and subclasses.

In Section 4.5 and 4.6, we will present the mean values of *deft* and the coefficients of variation by variable category, along with sampling errors and design effects of the selected variables. The complete list of sampling errors and design effects for all the 60 variables is attached in Appendix I.

**Table 4.8: Characteristics of Category and The Variables**

| Category | Description and Variable List |
|---|---|
| 1. Durable goods | This category consists of variables that measure the proportion of various durable goods owned by the household. |
| Variables | DURA1, DURA11, **DURA12**, DURA13, DURA23, DURA24, DURA25, DURA26, DURA27, DURA28, **DURA29**, DURA3, DURA30, DURA31, DURA32, DURA33, DURA34, **DURA35**, DURA36, DURA37, DURA4 and DURA5 |
| 2. Consumer Products | This category consists of variables that measure the proportion of various daily consumer products bought by the household in the past one month. |
| Variables | P1MBUY1, P1MBUY13, P1MBUY14, **P1MBUY15**, P1MBUY17, **P1MBUY18**, P1MBUY19, P1MBUY20, P1MBUY22, P1MBUY3, P1MBUY4, P1MBUY5, P1MBUY6 and P1MBUY9 |
| 3. Household General* | This category consists of variables that measure the mean of car ownership, TV set and member in the household. |
| Variables | CAR, **MAABV15**, MABLW15 and **TVSET** |
| 4. Media consumption | This category consists of variables that measure the proportion of respondents read newspaper and the mean of days they watched television in a week. |
| Variables | **FREQ1**, FREQ2, FREQ3, NPP7DA, NPP7DB, NPP7DE, **NPYTDA**, NPYTDB and NPYTDE |
| 5. Individual General | This category consists of variables that measure the proportion of respondents using certain banking facilities, literacy rate and leisure activities commonly participated. |
| Variables | **ACTVT1**, ACTVT2, ACTVT4, ACTVT5, ACTVT6, BANK1, **BANK13**, BANK15, BANK6, LITER1 and **LITER3** |

* The variables CAR and TVSET measure ownership of durable goods. However, these two variables are grouped into Household General category because they are measured in ratio scale.

## 4.4 The Patterns of Design Effects (*deft*) for Media Index Survey (MIS)

For the total sample, the overall *deft* averaged over all variables is 1.77. This means that due to the complex sample design adopted in MIS, the variance is increased, on average by a factor of $deft^2 = 3.13$ over that in an equivalent SRS design. Compared to the average *deft* value of 1.61 obtained in the 1996 Family Planning Acceptors Survey (Kish, et al., 1976), the average *deft* value for MIS is about 10% higher. The major difference between these two surveys is the sampling design adopted. In the 1996 Family Planning Acceptors Survey, the sample for metropolitan and urban strata were drawn from individual women which gave the average *deft* value of approximately 1. For the rural stratum, cluster sampling was employed with the selection of mukims (administrative units) at average of 70.86 women per mukim and yielded average *deft* value of 1.92. In MIS, a uniform design was adopted for all strata, yielding approximately equal average *deft* value for all strata (see Table 4.11).

The overall *deft* value of 1.77 indicates that due to complex sample design adopted in MIS, the effective sample size of the survey is equivalent to SRS design of 2792 (8738 divide by 3.13). Failure to incorporate the design effects in the confidence interval statements will distort the accuracy of the estimates. For example, if one desires to make correct statements at 95% confidence level (p=0.95, $t_\alpha = 2$), the actual confidence level of making correct statements is 74% due to the design effects, corresponding to $t'_\alpha = (2 / 1.77) = 1.13$. In other words, the probability of making an erroneous conclusion has increased from 5% to 26% if SRS is assumed.

Table 4.9 shows mean values of *deft* and coefficients of variation for total sample and Variable Category. The average *deft* values vary from 1.60 for Household General category to 1.98 for Media Consumption. Media Consumption category has the highest average *deft* value because variables within the category have large *deft* values between 1.71 to 2.36 with *deft* values above 2.0 for three main national TV channels (see Table 4.10 - FREQ1, FREQ2 and FREQ3).

The *deft* value for variable FREQ1 is 2.36, which is the largest in the category. This will enlarge the 95% confidence interval from (4.37 ± 0.066 days) to (4.37 ± 0.156 days), or reduce the confidence level from 95% assuming SRS to 60% for MIS sampling design, for the same width of the confidence interval. One of the factors contributing to this situation is the great difference among ethnic groups on the average days in which respondents turned on TV1 programs per week. The clustering effect of population notably by ethnicity has a great impact on this variable.

On the other hand, the proportion of respondents who have read any newspapers the day before the survey (NPYTDA) has the *deft* value of 1.80, which is larger than the overall average of 1.77. Exploratory data analysis shows that the mean value of newspaper readership varies widely within stratum and ethnicity, though to a lesser extent compared to FREQ1 variable. This is attributed to the reduction of the clustering impact by ethnicity as compared to FREQ1. On average, the variance in Media Consumption category is increased by a factor of 3.92 for the MIS design adopted.

Durable Goods and Consumer Products categories have about the same average *deft* value. The average *deft* value is 1.75 with coefficient of variation of

for Durable Goods category, while the *deft* value for Consumer Products is with coefficient of variation of 13%. Contrary to expectation, the *deft* values in categories fluctuate without any clear pattern in relation to the penetration rate durable goods or daily products (see Table 4.10). For example, about 82% 88% of households own electric rice cooker (DURA27) and refrigerator RA29) with *deft* values of 1.63 and 1.60, respectively. On the other hand, A1 (portable radio/cassette player) and DURA33 (portable fan) with ration rate of 89% and 83% respectively have *deft* values of 1.82 and 1.94 ctively. This may be due to the uneven ownership rate of certain durable goods s the PSUs.

The *deft* value is 2.08 for air conditioner ownership (DURA35), which is that its variance will be increased by 4.33 times for the MIS sampling design mpared to SRS design. The installation of air conditioner is closely related to of dwelling unit and the social economic status of the households. For instance, very rare to see wooden houses equipped with air conditioner. Condominiums double storey houses are much likely to be equipped with air conditioners pared to single storey terrace houses or low cost flats.

The Individual General category consisting of 11 variables has an average value of 1.68 with coefficient of variation of 12%. The *deft* value for each able in this category is relatively low, ranging from 1.47 to 1.76 except for TVT4 (Activity- travelling) with a *deft* value of 2.21. Comparison of *deft* values veen variable ACTVT1 (1.56) and LITER3 (1.77) shows greater variations in acy rate in Bahasa Malaysia as compared to participation in sports activities. average effective sample size for this category is equivalent to SRS design of 3096.

The smallest average *deft* value (1.60) is recorded for Household General category but it has the biggest coefficient of variation (14%). Variables in the category are number of members aged 15 years old and above (MAABV15), number of members aged below 15 years old (MABLW15), number of TV sets (TVSET) and number of cars owned by the households (CAR). Among these variables, CAR has the largest *deft* value (1.91) because the number of cars owned by households varies from 0 to 5 cars. The *deft* value of 1.59 for TVSET will increase its 95% confidence interval from (1.12 ± 0.010) to (1.12 ± 0.016). Due to the complex sample design used in MIS, the range of 95% confidence interval increases by 59% over that obtained in simple random sampling. This would in turn translate the 95% confidence interval assumed in SRS to 80% confidence level based on the complex design used in the MIS.

In the overall sample, variable DURA5 has the smallest *deft* value (1.37) and variable FREQ1 has the greatest *deft* value (2.36). Therefore, the variances of variables would be increased by a factor between 1.88 to 5.57. In view of the variations of *deft* values among the variables, attention should also be given to the relative error of each variable (Table 4.10). For example, even though variable DURA35 has *deft* value of 2.08 but the relative error is 0.054 (coefficient of variation of 5.4%). Thus, if one is satisfied with 5.4% coefficient of variation then the MIS sampling design is still appropriate for variable DURA35.

**Table 4.9: Mean Values and Coefficients of Variation of Design Effects (*deft*) for Total Sample and Variable Category**

| Category | No of Variables | Mean of design effect (*deft*) | Standard deviation | Coefficient of variation |
|---|---|---|---|---|
| 1. Durable Goods | 22 | 1.7487 | 0.19 | 0.11 |
| 2. Consumer Products | 14 | 1.7874 | 0.23 | 0.13 |
| 3. Household General | 4 | 1.6007 | 0.22 | 0.14 |
| 4. Media Consumption | 9 | 1.9824 | 0.22 | 0.11 |
| 5. Individual General | 11 | 1.6842 | 0.19 | 0.12 |
| Total Sample | 60 | 1.7711 | 0.23 | 0.13 |

**Table 4.10: Sampling Errors - Entire Sample, MIS**

| Variable | Value | Std error | No of cases | Std error assuming SRS | Std deviation | Design effect | Relative error |
|---|---|---|---|---|---|---|---|
| | r | se | n | $se_{srs}$ | sd | *deft* | se/r |
| CAR | 0.6261 | 0.0148 | 8738 | 0.0077 | 0.7234 | 1.9149 | 0.0237 |
| DURA1 | 0.8909 | 0.0061 | 8738 | 0.0033 | 0.3117 | 1.8197 | 0.0068 |
| DURA11 | 0.3535 | 0.0083 | 8738 | 0.0051 | 0.4781 | 1.6283 | 0.0236 |
| DURA12 | 0.7249 | 0.0083 | 8738 | 0.0048 | 0.4466 | 1.7328 | 0.0114 |
| DURA13 | 0.1921 | 0.0071 | 8738 | 0.0042 | 0.3940 | 1.6917 | 0.0371 |
| DURA23 | 0.2882 | 0.0084 | 8738 | 0.0048 | 0.4529 | 1.7379 | 0.0292 |
| DURA24 | 0.1182 | 0.0056 | 8738 | 0.0035 | 0.3229 | 1.6129 | 0.0471 |
| DURA25 | 0.6958 | 0.0092 | 8738 | 0.0049 | 0.4601 | 1.8731 | 0.0132 |
| DURA26 | 0.2810 | 0.0084 | 8738 | 0.0048 | 0.4495 | 1.7548 | 0.0300 |
| DURA27 | 0.8168 | 0.0068 | 8738 | 0.0041 | 0.3869 | 1.6343 | 0.0083 |
| DURA28 | 0.3036 | 0.0091 | 8738 | 0.0049 | 0.4598 | 1.8457 | 0.0299 |
| DURA29 | 0.8816 | 0.0055 | 8738 | 0.0035 | 0.3232 | 1.6049 | 0.0063 |
| DURA3 | 0.2196 | 0.0070 | 8738 | 0.0044 | 0.4140 | 1.5883 | 0.0320 |
| DURA30 | 0.7216 | 0.0084 | 8738 | 0.0048 | 0.4483 | 1.7587 | 0.0117 |
| DURA31 | 0.2962 | 0.0089 | 8738 | 0.0049 | 0.4566 | 1.8245 | 0.0301 |
| DURA32 | 0.1801 | 0.0083 | 8738 | 0.0041 | 0.3843 | 2.0233 | 0.0462 |
| DURA33 | 0.8306 | 0.0078 | 8738 | 0.0040 | 0.3751 | 1.9401 | 0.0094 |
| DURA34 | 0.6750 | 0.0110 | 8738 | 0.0050 | 0.4684 | 2.1939 | 0.0163 |
| DURA35 | 0.1441 | 0.0078 | 8738 | 0.0038 | 0.3512 | 2.0780 | 0.0542 |
| DURA36 | 0.4482 | 0.0088 | 8738 | 0.0053 | 0.4973 | 1.6454 | 0.0195 |
| DURA37 | 0.1186 | 0.0057 | 8738 | 0.0035 | 0.3233 | 1.6533 | 0.0482 |
| DURA4 | 0.2551 | 0.0068 | 8738 | 0.0047 | 0.4359 | 1.4635 | 0.0268 |
| DURA5 | 0.1144 | 0.0047 | 8738 | 0.0034 | 0.3184 | 1.3663 | 0.0407 |
| MAABV15 | 3.0409 | 0.0236 | 8738 | 0.0158 | 1.4735 | 1.4968 | 0.0078 |
| MABLW15 | 1.5745 | 0.0238 | 8738 | 0.0170 | 1.5900 | 1.3972 | 0.0151 |
| P1MBUY1 | 0.5160 | 0.0088 | 8738 | 0.0053 | 0.4998 | 1.6446 | 0.0170 |
| P1MBUY13 | 0.1693 | 0.0077 | 8738 | 0.0040 | 0.3750 | 1.9095 | 0.0453 |
| P1MBUY14 | 0.1836 | 0.0082 | 8738 | 0.0041 | 0.3872 | 1.9915 | 0.0449 |
| P1MBUY15 | 0.4034 | 0.0104 | 8738 | 0.0052 | 0.4906 | 1.9811 | 0.0258 |
| P1MBUY17 | 0.2673 | 0.0089 | 8738 | 0.0047 | 0.4426 | 1.8799 | 0.0333 |

| Variable | Value | Std error | No of cases | Std error assuming SRS | Std deviation | Design effect | Relative error |
|---|---|---|---|---|---|---|---|
| | r | se | n | $se_{srs}$ | sd | *deft* | se/r |
| P1MBUY18 | 0.1192 | 0.0049 | 8738 | 0.0035 | 0.3241 | 1.4134 | 0.0411 |
| P1MBUY19 | 0.1012 | 0.0050 | 8738 | 0.0032 | 0.3016 | 1.5347 | 0.0489 |
| P1MBUY20 | 0.1721 | 0.0088 | 8738 | 0.0040 | 0.3775 | 2.1771 | 0.0511 |
| P1MBUY22 | 0.1276 | 0.0068 | 8738 | 0.0036 | 0.3337 | 1.9152 | 0.0536 |
| P1MBUY3 | 0.1052 | 0.0054 | 8738 | 0.0033 | 0.3068 | 1.6392 | 0.0512 |
| P1MBUY4 | 0.1733 | 0.0066 | 8738 | 0.0040 | 0.3785 | 1.6301 | 0.0381 |
| P1MBUY5 | 0.3994 | 0.0090 | 8738 | 0.0052 | 0.4898 | 1.7163 | 0.0225 |
| P1MBUY6 | 0.3416 | 0.0105 | 8738 | 0.0051 | 0.4743 | 2.0752 | 0.0308 |
| P1MBUY9 | 0.1094 | 0.0051 | 8738 | 0.0033 | 0.3122 | 1.5160 | 0.0463 |
| TVSET | 1.1214 | 0.0081 | 8738 | 0.0051 | 0.4768 | 1.5941 | 0.0073 |
| ACTVT1 | 0.2470 | 0.0072 | 8738 | 0.0046 | 0.4313 | 1.5633 | 0.0292 |
| ACTVT2 | 0.8352 | 0.0062 | 8738 | 0.0040 | 0.3710 | 1.5508 | 0.0074 |
| ACTVT4 | 0.4496 | 0.0118 | 8738 | 0.0053 | 0.4975 | 2.2081 | 0.0261 |
| ACTVT5 | 0.9025 | 0.0055 | 8738 | 0.0032 | 0.2967 | 1.7226 | 0.0061 |
| ACTVT6 | 0.9028 | 0.0047 | 8738 | 0.0032 | 0.2962 | 1.4737 | 0.0052 |
| BANK1 | 0.2861 | 0.0079 | 8738 | 0.0048 | 0.4520 | 1.6385 | 0.0277 |
| BANK13 | 0.4261 | 0.0089 | 8738 | 0.0053 | 0.4945 | 1.6909 | 0.0210 |
| BANK15 | 0.1147 | 0.0055 | 8738 | 0.0034 | 0.3186 | 1.6104 | 0.0479 |
| BANK6 | 0.7230 | 0.0076 | 8738 | 0.0048 | 0.4475 | 1.5789 | 0.0105 |
| FREQ1 | 4.3748 | 0.0779 | 8738 | 0.0330 | 3.0828 | 2.3625 | 0.0178 |
| FREQ2 | 5.2573 | 0.0569 | 8738 | 0.0281 | 2.6233 | 2.0291 | 0.0108 |
| FREQ3 | 5.7004 | 0.0543 | 8738 | 0.0259 | 2.4255 | 2.0933 | 0.0095 |
| LITER1 | 0.4812 | 0.0092 | 8738 | 0.0053 | 0.4997 | 1.7216 | 0.0191 |
| LITER3 | 0.7941 | 0.0076 | 8738 | 0.0043 | 0.4044 | 1.7671 | 0.0096 |
| NPP7DA | 0.7225 | 0.0082 | 8738 | 0.0048 | 0.4478 | 1.7091 | 0.0113 |
| NPP7DB | 0.4314 | 0.0118 | 8738 | 0.0053 | 0.4953 | 2.2180 | 0.0272 |
| NPP7DE | 0.1710 | 0.0073 | 8738 | 0.0040 | 0.3765 | 1.8239 | 0.0430 |
| NPYTDA | 0.5699 | 0.0096 | 8738 | 0.0053 | 0.4951 | 1.8033 | 0.0168 |
| NPYTDB | 0.2998 | 0.0098 | 8738 | 0.0049 | 0.4582 | 1.9935 | 0.0326 |
| NPYTDE | 0.1321 | 0.0066 | 8738 | 0.0036 | 0.3386 | 1.8086 | 0.0496 |

## 4.5 The Patterns of Design Effects by Stratum

Overall, there is only a small difference of average *deft* for all variables across strata. This is not surprising in view of the fact that similar sampling designs and approximately equal cluster sizes were used in the MIS within each stratum.

Table 4.11 shows the average values of *deft* and coefficients of variation by stratum and Variable Category. The UT stratum has the largest average *deft* value of 1.84 with coefficient of variation of 15%. In contrast, the RT stratum has the smallest average *deft* value of 1.69. The urban-rural differentials in *deft* may have something to do with the rural society structure that tends to be less complex and more homogenous as compared to urban strata (MT and UT). This can be further vindicated by the average *deft* values of each Variable Category, where the average *deft* value for each category in RT stratum is equal to or lower than 1.77, the total sample average *deft* value.

With the sampling design adopted in MIS, the effective sample sizes are equivalent to SRS design of size 921, 826 and 1035 respectively for MT, UT and RT. Therefore, the probability of making correct conclusion is reduced from 95% to 73%, 72% and 76% for MT, UT and RT respectively if SRS is assumed.

It is interesting to observe that the average *deft* values by Variable Category and stratum did not show similar trend or pattern. In the MT and UT strata, Media Consumption category has the highest average *deft* value within the stratum. For RT stratum the average *deft* value is greatest for Consumer Products category with average *deft* value of 1.78. In contrast, the lowest average *deft* value is the Individual General category for MT stratum, and for UT and RT it is Household

General category. This may be due to the uneven distribution of population with respects to the variables in each Variable Category.

Ownership of telephone (DURA12), refrigerator (DURA29) and air conditioner (DURA35) are the selected variables from Durable Goods category for analysis (Table 4.12). Not surprisingly, the RT stratum has the largest *deft* value for DURA12 (1.87) and DURA29 (1.70), but the *deft* values for both products in MT stratum are only slightly smaller than RT stratum. This shows that telephone and refrigerator ownership among the households in MT and RT strata are less randomly distributed across population compared to UT stratum. Anyway, the coefficients of variation for these two variables are below 3% in all strata.

Air conditioner ownership (DURA35), which is highly dependent on the type of houses and the household income, has the highest *deft* value in MT (2.34), followed by UT (1.94) and RT (1.65). In RT stratum, the type of houses tends to be more homogeneous over all areas and the variation of household income is also relatively smaller. This yields homogeneity not only within the PSUs but also between PSUs. The other factor contributing to the low design effect is the low level of air conditioner ownership (5%) in RT stratum. The high homogeneity across clusters in the rural areas can be observed from the low *deft* values for most of the variables compared to the same variables in MT and UT strata.

Investigation on the 2 daily consumer products purchased in the past 1 month shows that the *deft* value for detergent (P1MBUY15) ranges from 1.85 to 2.11 across strata. Thus, the sampling design in MIS causes the variance of P1MBUY15 to increase by about 4 times compared to an identical sample size of SRS design. The *deft* values for P1MBUY18 is between 1.37 and 1.47 across strata and its variance will increase by about 2 times for the MIS sample.

The *deft* values for MAABV15 and TVSET, which represent the mean values, are between 1.36 to 1.75 across strata with coefficients of variation of below 2%. However, the *deft* value for variable FREQ1 is significantly larger compared to MAABV15 and TVSET. The *deft* value for variable FREQ1 is 2.54 for MT, 2.66 for UT and 2.01 for RT. For instance, the 95% confidence interval for MT stratum is increased from (3.34 ± 0.116 days) to (3.34 ± 0.294 days). The large *deft* value for number of days respondents turned on TV1 programs in a week (FREQ1) especially in urban areas (MT and UT) may be explained by the fact that the urban population have greater access to a wider range of electronic media and other modern objects.

An inspection of the results indicates that, as the society becomes more complex with the more choices and marketing forces, the purchase of durable goods and daily consumer products, as well as the individual media consumption habits, tends to be more differentiated across the population. Two general trends can be observed from Table 4.12. First, the design effects in RT stratum tend to be smaller than those in the urban strata (MT and UT). Second, the coefficients of variation of statistics in RT stratum are likely to be greater than the urban strata.

**Table 4.11: Means and Coefficients of Variation of Design Effects (*deft*) by Stratum and Variable Category**

| Category | MT | | UT | | RT | |
|---|---|---|---|---|---|---|
| | Mean | *cv* | Mean | *cv* | Mean | *cv* |
| 1. Durable Goods | 1.8614 | 0.17 | 1.7515 | 0.12 | 1.6603 | 0.11 |
| 2. Consumer Products | 1.6984 | 0.12 | 1.9179 | 0.14 | 1.7818 | 0.17 |
| 3. Household General | 1.7947 | 0.17 | 1.6125 | 0.14 | 1.3777 | 0.07 |
| 4. Media Consumption | 2.0878 | 0.13 | 2.1242 | 0.14 | 1.7403 | 0.14 |
| 5. Individual General | 1.5939 | 0.13 | 1.7882 | 0.12 | 1.7178 | 0.15 |
| Total | 1.8039 | 0.17 | 1.8437 | 0.15 | 1.6924 | 0.15 |

**Table 4.12: Sampling Errors of Selected Variables - By Stratum, MIS**

| Variable | Value | Std error | No of cases | Std error assuming SRS | Std deviation | Design effect | Relative error |
|---|---|---|---|---|---|---|---|
| | r | se | n | $se_{srs}$ | sd | *deft* | se/r |
| **MT** | | | | | | | |
| DURA12 | 0.8412 | 0.0121 | 2985 | 0.0067 | 0.3655 | 1.8145 | 0.0144 |
| DURA29 | 0.9209 | 0.0080 | 2985 | 0.0049 | 0.2699 | 1.6166 | 0.0087 |
| DURA35 | 0.2590 | 0.0188 | 2985 | 0.0080 | 0.4381 | 2.3435 | 0.0726 |
| MAABV15 | 3.2114 | 0.0453 | 2985 | 0.0282 | 1.5418 | 1.6063 | 0.0141 |
| P1MBUY15 | 0.4023 | 0.0166 | 2985 | 0.0090 | 0.4905 | 1.8546 | 0.0414 |
| P1MBUY18 | 0.1414 | 0.0087 | 2985 | 0.0064 | 0.3485 | 1.3658 | 0.0616 |
| TVSET | 1.2291 | 0.0179 | 2985 | 0.0102 | 0.5568 | 1.7539 | 0.0145 |
| ACTVT1 | 0.3035 | 0.0140 | 2985 | 0.0084 | 0.4599 | 1.6627 | 0.0461 |
| BANK13 | 0.5638 | 0.0145 | 2985 | 0.0091 | 0.4960 | 1.5990 | 0.0257 |
| FREQ1 | 3.3430 | 0.1472 | 2985 | 0.0580 | 3.1686 | 2.5387 | 0.0440 |
| LITER3 | 0.8214 | 0.0116 | 2985 | 0.0070 | 0.3830 | 1.6535 | 0.0141 |
| NPYTDA | 0.7357 | 0.0153 | 2985 | 0.0081 | 0.4410 | 1.8931 | 0.0208 |
| **UT** | | | | | | | |
| DURA12 | 0.7673 | 0.0133 | 2798 | 0.0080 | 0.4226 | 1.6652 | 0.0173 |
| DURA29 | 0.8956 | 0.0085 | 2798 | 0.0058 | 0.3058 | 1.4757 | 0.0095 |
| DURA35 | 0.1308 | 0.0123 | 2798 | 0.0064 | 0.3373 | 1.9366 | 0.0944 |
| MAABV15 | 2.9339 | 0.0398 | 2798 | 0.0268 | 1.4151 | 1.4887 | 0.0136 |
| P1MBUY15 | 0.4271 | 0.0185 | 2798 | 0.0094 | 0.4947 | 1.9750 | 0.0433 |
| P1MBUY18 | 0.1108 | 0.0085 | 2798 | 0.0059 | 0.3139 | 1.4260 | 0.0764 |
| TVSET | 1.1122 | 0.0129 | 2798 | 0.0081 | 0.4298 | 1.5929 | 0.0116 |
| ACTVT1 | 0.2430 | 0.0129 | 2798 | 0.0081 | 0.4290 | 1.5858 | 0.0529 |
| BANK13 | 0.4199 | 0.0179 | 2798 | 0.0093 | 0.4936 | 1.9218 | 0.0427 |
| FREQ1 | 4.6719 | 0.1512 | 2798 | 0.0568 | 3.0037 | 2.6627 | 0.0324 |
| LITER3 | 0.7888 | 0.0147 | 2798 | 0.0077 | 0.4082 | 1.9051 | 0.0186 |
| NPYTDA | 0.5976 | 0.0194 | 2798 | 0.0093 | 0.4905 | 2.0964 | 0.0325 |
| **RT** | | | | | | | |
| DURA12 | 0.5672 | 0.0170 | 2955 | 0.0091 | 0.4956 | 1.8686 | 0.0300 |
| DURA29 | 0.8284 | 0.0118 | 2955 | 0.0069 | 0.3771 | 1.6988 | 0.0142 |
| DURA35 | 0.0406 | 0.0060 | 2955 | 0.0036 | 0.1974 | 1.6488 | 0.1474 |
| MAABV15 | 2.9699 | 0.0367 | 2955 | 0.0265 | 1.4420 | 1.3846 | 0.0124 |
| P1MBUY15 | 0.3821 | 0.0189 | 2955 | 0.0089 | 0.4860 | 2.1122 | 0.0494 |
| P1MBUY18 | 0.1049 | 0.0083 | 2955 | 0.0056 | 0.3065 | 1.4663 | 0.0788 |
| TVSET | 1.0213 | 0.0101 | 2955 | 0.0074 | 0.4030 | 1.3612 | 0.0099 |
| ACTVT1 | 0.1936 | 0.0103 | 2955 | 0.0073 | 0.3952 | 1.4212 | 0.0534 |
| BANK13 | 0.2927 | 0.0140 | 2955 | 0.0084 | 0.4551 | 1.6736 | 0.0479 |
| FREQ1 | 5.1357 | 0.1023 | 2955 | 0.0510 | 2.7725 | 2.0058 | 0.0199 |
| LITER3 | 0.7716 | 0.0134 | 2955 | 0.0077 | 0.4199 | 1.7367 | 0.0174 |
| NPYTDA | 0.3763 | 0.0149 | 2955 | 0.0089 | 0.4845 | 1.6664 | 0.0395 |

**Table 4.13: ANOVA on Means of Design Effects (*deft*) by Stratum and Variable Category**

| | | | Unique Method | | | | |
|---|---|---|---|---|---|---|---|
| | | | Sum of Squares | df | Mean Square | F | Sig. |
| DEFT | Main Effects | (Combined) | .343 | 6 | 5.718E-02 | 3.145 | .069 |
| | | CATEGORY | .247 | 4 | 6.177E-02 | 3.398 | .066 |
| | | STRATUM | 9.599E-02 | 2 | 4.799E-02 | 2.640 | .132 |
| | Model | | .343 | 6 | 5.718E-02 | 3.145 | .069 |
| | Residual | | .145 | 8 | 1.818E-02 | | |
| | Total | | .488 | 14 | 3.489E-02 | | |

**Table 4.14: ANOVA on Design Effects (*deft*) by Stratum and Selected Variables, MIS**

| | | | Unique Method | | | | |
|---|---|---|---|---|---|---|---|
| | | | Sum of Squares | df | Mean Square | F | Sig. |
| DEFT | Main Effects | (Combined) | 2.605 | 13 | .200 | 6.012 | .000 |
| | | Selected Variables | 2.450 | 11 | .223 | 6.681 | .000 |
| | | STRATUM | .156 | 2 | 7.775E-02 | 2.333 | .121 |
| | Model | | 2.605 | 13 | .200 | 6.012 | .000 |
| | Residual | | .733 | 22 | 3.333E-02 | | |
| | Total | | 3.339 | 35 | 9.539E-02 | | |

Analysis of variance (ANOVA) is performed to examine the effect of geographical domains, Variable Category and selected variables on the *deft* values. Table 4.13 shows that the average *deft* values are not significantly different across geographical domains (F=2.640, p = 0.132), which indicates that the design effect is rather similar in all strata. However, at 0.10 level of significance, Variable Category has a significant effect on the average *deft* value (F=3.398, p = 0.066). The differences in *deft* value among sample statistics are more pronounced in the analysis of selected variables (Table 4.14) where the results show that *deft* value is significantly different across selected variables (F=6.681, p = 0.000).

## 4.6 The Patterns of Design Effects for Subclasses

Surveys always involve numerous subclasses of interest. The subclasses are very important sub-populations that are actually different from one another in many characteristics. For example, in a multi-racial country like Malaysia, subclass analysis by ethnic groups is vital for a better understanding of population. Therefore, it is necessary to develop adequate models for the patterns of variation of sampling errors for subclasses. Design effects of subclasses are usually defined in terms of demographic and social-economic characteristics of individual respondents. The subclasses considered in this section are ethnic group, gender and age group of respondents. The complete list of design effects for total sample by above subclasses is presented in Appendix I.

Normally, these subclasses do not form explicit domains for sampling design and selection. They are also generally spread across the population and sometimes distributed throughout the sample clusters. But, they may or may not be randomly distributed across the sample clusters. Therefore, the *deft* values are quite different because smaller effective cluster size tends to reduce the design effects of subclasses.

### 4.6.1 The Patterns of Design Effects by Ethnic Group

As expected, the average values of *deft* for all variables by ethnic group are smaller than the average *deft* value for total sample. Table 4.15 illustrates the mean values of *deft* and the coefficients of variation by ethnic group and Variable Category. The average *deft* value is 1.62 for Malays, 1.47 for Chinese and 1.21 for Indians/Others, with the Chinese having the largest coefficient of variation (14%). The average *deft* value for each of the Variable Category does not deviate much

from the total average *deft* value, across ethnic groups. In other words, the average *deft* values of every Variable Category in Malay subclass are larger than the category in Chinese subclass, whereas the Indians/Others subclass has the smallest average *deft* values.

Because of the complex design employed in MIS, the effective sample size by ethnic group is equivalent to SRS sample size of 1980, 1173 and 688 respectively for Malays, Chinese and Indians/Others. If one calculates the confidence interval without properly accounting for the design effect, one will increase the error rate significantly. For example, the error rate is increased from 5% to 22%, 17% and 10% for Malays, Chinese and Indians/Others respectively if SRS is assumed.

Both Malays and Chinese have the lower mean *deft* value for Household General category when compared to other categories within the ethnic group. For the Malays, the low average *deft* value for this category is due to the small *deft* value for variable MAABV15 (1.37) and MABLW15 (1.31), indicating that the Malay households tend to have more homogeneous in household size across geographical domains.

Category with highest average *deft* value among Chinese is Media Consumption category. The *deft* value for number of days turned on TV1 programs in a week (FREQ1) is 2.05 with coefficient of variation of about 5% (see Table 4.16). Large *deft* value of FREQ1 for Chinese is partly caused by the availability of cable and satellite television and TV programs from neighboring countries, which further differentiates the viewing habits from one to another areas. This will enlarge the 95% confidence interval from (2.14 ± 0.116 days) to (2.14 ± 0.237 days). The *deft* value for NPYTDA is small for Chinese because of the high proportion of

Chinese respondents who read any newspapers the day prior to the survey over all strata.

Among the selected variables in Table 4.16, the variables with higher *deft* values for Indians/Others are P1MBUY15 and FREQ1 with about the same *deft* value of 1.36. Thus, the probability of making correct conclusion is reduced from 95% to 86% if SRS is assumed.

The *deft* values on Consumer Products and Media Consumption categories are likely to be higher for Malays. The coefficients of variation are generally larger for selected variables in Indian/Others subclass. Further analysis using the coefficient of intra-class correlation is carried out in subsequent section to assess the clustering effect on the subclass because *deft* value depends on effective size of subclass.

**Table 4.15: Mean Values and Coefficients of Variation of Design Effects (*deft*) by Ethnic Group and Variable Category**

| Category | Malays | | Chinese | | Indians/Others | |
|---|---|---|---|---|---|---|
| | Mean | *cv* | Mean | *cv* | Mean | *cv* |
| 1. Durable Goods | 1.5708 | 0.10 | 1.4857 | 0.12 | 1.2174 | 0.10 |
| 2. Consumer Products | 1.7250 | 0.15 | 1.4435 | 0.11 | 1.2257 | 0.10 |
| 3. Household General | 1.4279 | 0.09 | 1.3933 | 0.11 | 1.2237 | 0.09 |
| 4. Media Consumption | 1.6942 | 0.06 | 1.5364 | 0.23 | 1.2340 | 0.08 |
| 5. Individual General | 1.6120 | 0.14 | 1.4230 | 0.11 | 1.1346 | 0.09 |
| Total | 1.6233 | 0.12 | 1.4658 | 0.14 | 1.2071 | 0.10 |

## Table 4.16: Sampling Errors of Selected Variables - By Ethnic Group, MIS

| Variable | Value | Std error | No of cases | Std error assuming SRS | Std deviation | Design effect | Relative error |
|---|---|---|---|---|---|---|---|
| | r | se | n | $se_{srs}$ | sd | *deft* | se/r |
| **Malays** | | | | | | | |
| DURA12 | 0.6507 | 0.0111 | 5196 | 0.0066 | 0.4768 | 1.6852 | 0.0171 |
| DURA29 | 0.8591 | 0.0075 | 5196 | 0.0048 | 0.3479 | 1.5485 | 0.0087 |
| DURA35 | 0.0439 | 0.0044 | 5196 | 0.0028 | 0.2048 | 1.5650 | 0.1014 |
| MAABV15 | 2.9149 | 0.0265 | 5196 | 0.0193 | 1.3948 | 1.3694 | 0.0091 |
| P1MBUY15 | 0.4167 | 0.0135 | 5196 | 0.0068 | 0.4931 | 1.9674 | 0.0323 |
| P1MBUY18 | 0.1386 | 0.0065 | 5196 | 0.0048 | 0.3455 | 1.3653 | 0.0472 |
| TVSET | 1.0916 | 0.0089 | 5196 | 0.0063 | 0.4522 | 1.4210 | 0.0082 |
| ACTVT1 | 0.2325 | 0.0084 | 5196 | 0.0059 | 0.4225 | 1.4414 | 0.0363 |
| BANK13 | 0.4365 | 0.0115 | 5196 | 0.0069 | 0.4960 | 1.6758 | 0.0264 |
| FREQ1 | 5.6166 | 0.0556 | 5196 | 0.0336 | 2.4200 | 1.6553 | 0.0099 |
| LITER3 | 0.8878 | 0.0065 | 5196 | 0.0044 | 0.3156 | 1.4782 | 0.0073 |
| NPYTDA | 0.4873 | 0.0119 | 5196 | 0.0069 | 0.4999 | 1.7166 | 0.0244 |
| **Chinese** | | | | | | | |
| DURA12 | 0.8750 | 0.0071 | 2535 | 0.0066 | 0.3308 | 1.0862 | 0.0082 |
| DURA29 | 0.9290 | 0.0070 | 2535 | 0.0051 | 0.2569 | 1.3766 | 0.0076 |
| DURA35 | 0.3657 | 0.0149 | 2535 | 0.0096 | 0.4817 | 1.5609 | 0.0408 |
| MAABV15 | 3.1795 | 0.0452 | 2535 | 0.0304 | 1.5284 | 1.4883 | 0.0142 |
| P1MBUY15 | 0.3815 | 0.0147 | 2535 | 0.0096 | 0.4858 | 1.5226 | 0.0385 |
| P1MBUY18 | 0.0832 | 0.0060 | 2535 | 0.0055 | 0.2763 | 1.0880 | 0.0717 |
| TVSET | 1.2071 | 0.0153 | 2535 | 0.0108 | 0.5434 | 1.4207 | 0.0127 |
| ACTVT1 | 0.2809 | 0.0131 | 2535 | 0.0089 | 0.4495 | 1.4649 | 0.0466 |
| BANK13 | 0.4118 | 0.0137 | 2535 | 0.0098 | 0.4923 | 1.4034 | 0.0333 |
| FREQ1 | 2.1365 | 0.1183 | 2535 | 0.0578 | 2.9081 | 2.0473 | 0.0553 |
| LITER3 | 0.6237 | 0.0139 | 2535 | 0.0096 | 0.4846 | 1.4431 | 0.0223 |
| NPYTDA | 0.7515 | 0.0109 | 2535 | 0.0086 | 0.4322 | 1.2657 | 0.0145 |
| **Indians/Others** | | | | | | | |
| DURA12 | 0.7299 | 0.0177 | 1007 | 0.0140 | 0.4442 | 1.2658 | 0.0243 |
| DURA29 | 0.8779 | 0.0124 | 1007 | 0.0103 | 0.3276 | 1.1992 | 0.0141 |
| DURA35 | 0.1033 | 0.0126 | 1007 | 0.0096 | 0.3045 | 1.3109 | 0.1218 |
| MAABV15 | 3.3416 | 0.0598 | 1007 | 0.0518 | 1.6451 | 1.1529 | 0.0179 |
| P1MBUY15 | 0.3903 | 0.0210 | 1007 | 0.0154 | 0.4881 | 1.3626 | 0.0537 |
| P1MBUY18 | 0.1102 | 0.0102 | 1007 | 0.0099 | 0.3133 | 1.0288 | 0.0922 |
| TVSET | 1.0596 | 0.0146 | 1007 | 0.0121 | 0.3841 | 1.2047 | 0.0138 |
| ACTVT1 | 0.2363 | 0.0144 | 1007 | 0.0134 | 0.4250 | 1.0765 | 0.0610 |
| BANK13 | 0.4081 | 0.0168 | 1007 | 0.0155 | 0.4917 | 1.0830 | 0.0411 |
| FREQ1 | 3.6018 | 0.1335 | 1007 | 0.0987 | 3.1315 | 1.3526 | 0.0371 |
| LITER3 | 0.7398 | 0.0146 | 1007 | 0.0138 | 0.4390 | 1.0526 | 0.0197 |
| NPYTDA | 0.5392 | 0.0203 | 1007 | 0.0157 | 0.4987 | 1.2920 | 0.0377 |

**Table 4.17: ANOVA on Means of Design Effects (*deft*) by Ethnic Group and Variable Category**

| | | | Unique Method | | | | |
|---|---|---|---|---|---|---|---|
| | | | Sum of Squares | df | Mean Square | F | Sig. |
| DEFT | Main Effects | (Combined) | .444 | 6 | 7.402E-02 | 16.382 | .000 |
| | | CATEGORY | 3.802E-02 | 4 | 9.505E-03 | 2.104 | .172 |
| | | RACE | .406 | 2 | .203 | 44.940 | .000 |
| | Model | | .444 | 6 | 7.402E-02 | 16.382 | .000 |
| | Residual | | 3.614E-02 | 8 | 4.518E-03 | | |
| | Total | | .480 | 14 | 3.430E-02 | | |

**Table 4.18: ANOVA on Design Effects (*deft*) by Ethnic Group and Selected Variables, MIS**

| | | | Unique Method | | | | |
|---|---|---|---|---|---|---|---|
| | | | Sum of Squares | df | Mean Square | F | Sig. |
| DEFT | Main Effects | (Combined) | 1.504 | 13 | .116 | 4.773 | .001 |
| | | RACE | .862 | 2 | .431 | 17.787 | .000 |
| | | Selected Variables | .642 | 11 | 5.835E-02 | 2.407 | .038 |
| | Model | | 1.504 | 13 | .116 | 4.773 | .001 |
| | Residual | | .533 | 22 | 2.424E-02 | | |
| | Total | | 2.037 | 35 | 5.821E-02 | | |

The ANOVA results in Table 4.17 strongly suggest that the average *deft* values are not the same among the ethnic groups (F=44.940, p = 0.000). At 0.05 significance level, the *deft* values among Variable Category are not significantly different (F=2.104, p = 0.172). However, at the same level of significance, the *deft* values for the selected variables are significantly different (Table 4.18, F=2.407, p = 0.038). This may suggest that the *deft* values indeed vary greatly among variables, but, when the variables are grouped into respective Variable Category, the average *deft* values are not significantly different among each other.

### 4.6.2 The Patterns of Design Effects by Gender

Gender is an important variable in most social survey or marketing research. The male-female differentials in the study variables are often of much relevance and interest to policy makers and researchers.

Overall, the average *deft* of Female subclass is slightly greater than Male subclass (1.51 versus 1.44) with the same coefficient of variation (11%). The effective sample size by gender is equivalent to SRS sample size of 1887 for Males and 2117 for Females. On average, the variance by gender subclass will increase by a factor of 2.28 for Males and 2.07 for Females. Further investigation by Variable Category shows that the average *deft* value for every category in Male subclass is smaller than its corresponding value of the same category in Female subclass. Table 4.19 shows the mean values of *deft* and the coefficients of variation by gender and Variable Category.

The coefficients of variation for Household variable categories (Durable Goods, Consumer Products and Household General) do not vary much between the two sexes. However, the coefficient of variation for the Males differs considerably from the Females with respect to Individual variable (Media Consumption and Individual General). Like in Media Consumption category, the coefficient of variation is 13% for Males compared to 9% for Females.

Table 4.20 illustrates the sampling errors and design effects of the selected variables by gender. By comparing every *deft* value between Males and Females, the differences for Household Variables are rather small because both Males and Females distributions in the population should not have any significant clustering

effects by household. This, in turn, should not impact on the interviews regardless of the gender of respondents.

The *deft* values on 2 of the Individual variables differ slightly greatly between Males and Females. The *deft* value of variable LITER3 is 1.41 for Males as compared to 1.57 for Females. The variable NPYTDA has *deft* value of 1.33 for Males and 1.59 for Females. This indicates that the distribution of Females who are literate in Bahasa Malaysia and who are newspaper readers are not as evenly spread as compared to the Males. On the other hand, the differences among variables ACTVT1, BANK13 and FREQ1 are less obvious between Males and Females.

If one were to calculate the 95% confidence interval for variable NPYTDA for both sexes, the likelihood of making an erroneous conclusion with the complex design has increased from 5% in SRS to 13% and 21% for Males and Females, respectively, if the confidence interval is held constant. The 95% confidence interval of proportion of Male respondents reading any newspapers the day before the survey increases from (0.67±0.015) to (0.67±0.020). As for the Females, the confidence interval increases from (0.49±0.014) to (0.49±0.023).

**Table 4.19: Mean Values and Coefficients of Variation of Design Effects (*deft*) by Gender and Variable Category**

| Category | Male | | Female | |
|---|---|---|---|---|
| | Mean | *cv* | Mean | *cv* |
| 1. Durable Goods | 1.4051 | 0.09 | 1.5018 | 0.09 |
| 2. Consumer Products | 1.4580 | 0.10 | 1.5483 | 0.11 |
| 3. Household General | 1.3642 | 0.07 | 1.4094 | 0.09 |
| 4. Media Consumption | 1.5951 | 0.13 | 1.6373 | 0.09 |
| 5. Individual General | 1.4066 | 0.09 | 1.4367 | 0.12 |
| Total | 1.4435 | 0.11 | 1.5149 | 0.11 |

## Table 4.20: Sampling Errors of Selected Variables - By Gender, MIS

| Variable | Value | Std error | No of cases | Std error assuming SRS | Std deviation | Design effect | Relative error |
|---|---|---|---|---|---|---|---|
| | r | se | n | $se_{srs}$ | sd | *deft* | se/r |
| **Male** | | | | | | | |
| DURA12 | 0.7242 | 0.0095 | 3912 | 0.0071 | 0.4470 | 1.3262 | 0.0131 |
| DURA29 | 0.8750 | 0.0070 | 3912 | 0.0053 | 0.3308 | 1.3292 | 0.0080 |
| DURA35 | 0.1521 | 0.0097 | 3912 | 0.0057 | 0.3592 | 1.6817 | 0.0635 |
| MAABV15 | 3.0844 | 0.0316 | 3912 | 0.0234 | 1.4624 | 1.3514 | 0.0102 |
| P1MBUY15 | 0.3318 | 0.0121 | 3912 | 0.0075 | 0.4709 | 1.6107 | 0.0365 |
| P1MBUY18 | 0.0946 | 0.0061 | 3912 | 0.0047 | 0.2927 | 1.3019 | 0.0644 |
| TVSET | 1.1378 | 0.0107 | 3912 | 0.0081 | 0.5048 | 1.3270 | 0.0094 |
| ACTVT1 | 0.3604 | 0.0103 | 3912 | 0.0077 | 0.4802 | 1.3362 | 0.0285 |
| BANK13 | 0.5153 | 0.0118 | 3912 | 0.0080 | 0.4998 | 1.4750 | 0.0229 |
| FREQ1 | 4.3078 | 0.0911 | 3912 | 0.0493 | 3.0859 | 1.8458 | 0.0211 |
| LITER3 | 0.8482 | 0.0081 | 3912 | 0.0057 | 0.3589 | 1.4099 | 0.0095 |
| NPYTDA | 0.6687 | 0.0100 | 3912 | 0.0075 | 0.4707 | 1.3294 | 0.0150 |
| **Female** | | | | | | | |
| DURA12 | 0.7254 | 0.0096 | 4826 | 0.0064 | 0.4463 | 1.4988 | 0.0133 |
| DURA29 | 0.8869 | 0.0066 | 4826 | 0.0046 | 0.3168 | 1.4424 | 0.0074 |
| DURA35 | 0.1376 | 0.0082 | 4826 | 0.0050 | 0.3445 | 1.6625 | 0.0599 |
| MAABV15 | 3.0056 | 0.0290 | 4826 | 0.0213 | 1.4817 | 1.3578 | 0.0096 |
| P1MBUY15 | 0.4615 | 0.0122 | 4826 | 0.0072 | 0.4986 | 1.6962 | 0.0264 |
| P1MBUY18 | 0.1392 | 0.0067 | 4826 | 0.0050 | 0.3462 | 1.3443 | 0.0481 |
| TVSET | 1.1082 | 0.0089 | 4826 | 0.0065 | 0.4524 | 1.3711 | 0.0081 |
| ACTVT1 | 0.1550 | 0.0069 | 4826 | 0.0052 | 0.3619 | 1.3273 | 0.0446 |
| BANK13 | 0.3537 | 0.0096 | 4826 | 0.0069 | 0.4782 | 1.3999 | 0.0272 |
| FREQ1 | 4.4291 | 0.0839 | 4826 | 0.0443 | 3.0796 | 1.8919 | 0.0189 |
| LITER3 | 0.7503 | 0.0098 | 4826 | 0.0062 | 0.4329 | 1.5772 | 0.0131 |
| NPYTDA | 0.4898 | 0.0114 | 4826 | 0.0072 | 0.4999 | 1.5879 | 0.0233 |

## Table 4.21: ANOVA on Means of Design Effects (*deft*) by Gender and Variable Category

| | | | Unique Method | | | | |
|---|---|---|---|---|---|---|---|
| | | | Sum of Squares | df | Mean Square | F | Sig. |
| DEFT | Main Effects | (Combined) | 7.290E-02 | 5 | 1.458E-02 | 31.601 | .003 |
| | | CATEGORY | 6.362E-02 | 4 | 1.591E-02 | 34.477 | .002 |
| | | SEX | 9.272E-03 | 1 | 9.272E-03 | 20.097 | .011 |
| | Model | | 7.290E-02 | 5 | 1.458E-02 | 31.601 | .003 |
| | Residual | | 1.845E-03 | 4 | 4.614E-04 | | |
| | Total | | 7.474E-02 | 9 | 8.305E-03 | | |

**Table 4.22: ANOVA on Design Effects (*deft*) by Gender and Selected Variables, MIS**

| | | | Unique Method | | | | |
|---|---|---|---|---|---|---|---|
| | | | Sum of Squares | df | Mean Square | F | Sig. |
| DEFT | Main Effects | (Combined) | .656 | 12 | 5.466E-02 | 12.157 | .000 |
| | | Selected Variables | .627 | 11 | 5.700E-02 | 12.678 | .000 |
| | | SEX | 2.891E-02 | 1 | 2.891E-02 | 6.429 | .028 |
| | Model | | .656 | 12 | 5.466E-02 | 12.157 | .000 |
| | Residual | | 4.946E-02 | 11 | 4.496E-03 | | |
| | Total | | .705 | 23 | 3.067E-02 | | |

The results in Table 4.21 suggest that the average *deft* values are significantly different between the two sexes (F=20.097, p = 0.011) and among Variable Categories (F=34.477, p = 0.002). The analysis by selected variables (Table 4.22) also gives the same conclusion. The results are in line with earlier observation that the *deft* values are not equal between Males and Females especially in Individual variable categories.

### 4.6.3 The Patterns of Design Effects by Age Group

Besides ethnicity and gender, age group is another important variable in most surveys. A person's needs and expectation change over his or her life cycle. An understanding of the age differentials with respect to the study variable is vital in identifying the target groups for programme intervention and formulation of marketing strategies. In the MIS, the respondents were grouped into 3 age groups, 15-29 years old, 30-49 years old and 50 years old and above. This grouping provides approximately even sample split to facilitate further analysis within the age group by stratum. Table 4.23 illustrates the mean values and coefficients of variation of *deft* by age group and Variable Category.

The average *deft* is 1.32 with coefficient of variation of 10% for age group 15-29 years old. The sampling design adopted in MIS corresponds to an effective sample size of 1547 for the SRS design. Household General and Individual General categories in this subclass have smaller average *deft* value of 1.23 and 1.24 respectively, but the coefficient of variation is the highest (13%) for Individual General among the categories. Table 4.24 illustrates the sampling errors and design effects for the selected variables by age group. Both variables MAABV15 and TVSET have the *deft* value of 1.23 and coefficient of variation of around 1%.

In Individual General category, the literacy rate in Bahasa Malaysia (LITER3) has a very small *deft* value of 1.14 with less than 0.5% coefficient of variation. This is because majority of the population (95%) in age group 15-29 years old are literate in Bahasa Malaysia. However, the large average *deft* value for Individual General category may be explained by the fact that while about 48% (ACTVT4) of respondents aged 15-29 years old claimed that travelling is one common leisure activity, significant variation exists across strata. For example, the proportion of respondents in MT stratum reported travelling as one common leisure activity is 55% with *deft* value of 1.51; the corresponding proportion for UT and RT strata is 46% and 41% respectively, with similar *deft* value of 1.65.

Media Consumption category has the highest average *deft* value of 1.42 among categories of the age group 15-29 years old subclass. Part of the contributions is from variable FREQ1 with *deft* value of 1.61. The *deft* value of 1.61 for FREQ1 will increase the error rate from 5% to 22% if one calculates the 95% confidence interval without taking into account the design effect of MIS sample.

Both age groups 30-49 and 50+ years old have the smallest average *deft* for Household General category, 1.34 and 1.14 respectively. The Media Consumption

category has the largest average *deft* in all the age groups. Contrary to expectation, respondents aged 50 years and older spent significantly less time watching television programs and reading newspapers as compared to the other 2 age groups. The average *deft* value for age group 50+ years old is lowest among age groups for the total sample and also in each Variable Category, primary due to the small effective sample size within the clusters. These noticeable differences imply the existence of differential effects by age group within category. Further analysis of coefficient of intra-class correlation are carried in the next section to assess the design effect to ensure the phenomena observed from *deft* values are genuine or due to differences in effective sample size within the clusters.

**Table 4.23: Mean Values and Coefficients of Variation of Design Effects (*deft*) by Age Group and Variable Category**

| Category | 15-29 | | 30-49 | | 50+ | |
|---|---|---|---|---|---|---|
| | Mean | *cv* | Mean | *Cv* | Mean | *cv* |
| 1. Durable Goods | 1.3164 | 0.08 | 1.4324 | 0.09 | 1.2470 | 0.07 |
| 2. Consumer Products | 1.3385 | 0.07 | 1.4928 | 0.10 | 1.2479 | 0.11 |
| 3. Household General | 1.2253 | 0.08 | 1.3408 | 0.12 | 1.1393 | 0.06 |
| 4. Media Consumption | 1.4219 | 0.10 | 1.5980 | 0.08 | 1.3419 | 0.07 |
| 5. Individual General | 1.2363 | 0.13 | 1.4657 | 0.09 | 1.2775 | 0.07 |
| Total | 1.3166 | 0.10 | 1.4713 | 0.10 | 1.2599 | 0.09 |

## Table 4.24: Sampling Errors of Selected Variables - By Age Group, MIS

| Variable | Value | Std error | No of cases | Std error assuming SRS | Std deviation | Design effect | Relative error |
|---|---|---|---|---|---|---|---|
| | r | se | n | $se_{srs}$ | sd | *deft* | se/r |
| **15-29 years** | | | | | | | |
| DURA12 | 0.7114 | 0.0115 | 2696 | 0.0087 | 0.4532 | 1.3171 | 0.0162 |
| DURA29 | 0.8657 | 0.0085 | 2696 | 0.0066 | 0.3410 | 1.2871 | 0.0098 |
| DURA35 | 0.1291 | 0.0089 | 2696 | 0.0065 | 0.3354 | 1.3850 | 0.0693 |
| MAABV15 | 3.6480 | 0.0385 | 2696 | 0.0314 | 1.6296 | 1.2269 | 0.0106 |
| P1MBUY15 | 0.3190 | 0.0125 | 2696 | 0.0090 | 0.4662 | 1.3961 | 0.0393 |
| P1MBUY18 | 0.1269 | 0.0077 | 2696 | 0.0064 | 0.3329 | 1.2035 | 0.0608 |
| TVSET | 1.1469 | 0.0122 | 2696 | 0.0099 | 0.5139 | 1.2313 | 0.0106 |
| ACTVT1 | 0.4458 | 0.0117 | 2696 | 0.0096 | 0.4972 | 1.2249 | 0.0263 |
| BANK13 | 0.4837 | 0.0123 | 2696 | 0.0096 | 0.4998 | 1.2729 | 0.0253 |
| FREQ1 | 4.4091 | 0.0952 | 2696 | 0.0590 | 3.0632 | 1.6136 | 0.0216 |
| LITER3 | 0.9592 | 0.0043 | 2696 | 0.0038 | 0.1979 | 1.1376 | 0.0045 |
| NPYTDA | 0.6306 | 0.0124 | 2696 | 0.0093 | 0.4827 | 1.3304 | 0.0196 |
| **30-49 years** | | | | | | | |
| DURA12 | 0.7413 | 0.0104 | 3861 | 0.0070 | 0.4380 | 1.4812 | 0.0141 |
| DURA29 | 0.9096 | 0.0057 | 3861 | 0.0046 | 0.2868 | 1.2408 | 0.0063 |
| DURA35 | 0.1585 | 0.0098 | 3861 | 0.0059 | 0.3653 | 1.6738 | 0.0621 |
| MAABV15 | 2.6768 | 0.0247 | 3861 | 0.0197 | 1.2221 | 1.2544 | 0.0092 |
| P1MBUY15 | 0.5009 | 0.0128 | 3861 | 0.0080 | 0.5001 | 1.5878 | 0.0255 |
| P1MBUY18 | 0.1676 | 0.0075 | 3861 | 0.0060 | 0.3735 | 1.2554 | 0.0450 |
| TVSET | 1.1365 | 0.0098 | 3861 | 0.0073 | 0.4527 | 1.3493 | 0.0086 |
| ACTVT1 | 0.1974 | 0.0091 | 3861 | 0.0064 | 0.3981 | 1.4241 | 0.0462 |
| BANK13 | 0.5131 | 0.0121 | 3861 | 0.0080 | 0.4999 | 1.4982 | 0.0235 |
| FREQ1 | 4.5405 | 0.0874 | 3861 | 0.0491 | 3.0514 | 1.7802 | 0.0193 |
| LITER3 | 0.8467 | 0.0089 | 3861 | 0.0058 | 0.3604 | 1.5327 | 0.0105 |
| NPYTDA | 0.6281 | 0.0114 | 3861 | 0.0078 | 0.4834 | 1.4669 | 0.0182 |
| **50+ years** | | | | | | | |
| DURA12 | 0.7125 | 0.0122 | 2181 | 0.0097 | 0.4527 | 1.2576 | 0.0171 |
| DURA29 | 0.8514 | 0.0099 | 2181 | 0.0076 | 0.3557 | 1.2987 | 0.0116 |
| DURA35 | 0.1371 | 0.0099 | 2181 | 0.0074 | 0.3440 | 1.3445 | 0.0722 |
| MAABV15 | 2.9349 | 0.0340 | 2181 | 0.0308 | 1.4397 | 1.1031 | 0.0116 |
| P1MBUY15 | 0.3352 | 0.0142 | 2181 | 0.0101 | 0.4722 | 1.4031 | 0.0423 |
| P1MBUY18 | 0.0243 | 0.0035 | 2181 | 0.0033 | 0.1540 | 1.0514 | 0.1427 |
| TVSET | 1.0633 | 0.0121 | 2181 | 0.0100 | 0.4659 | 1.2086 | 0.0113 |
| ACTVT1 | 0.0890 | 0.0078 | 2181 | 0.0061 | 0.2847 | 1.2851 | 0.0881 |
| BANK13 | 0.2008 | 0.0101 | 2181 | 0.0086 | 0.4007 | 1.1756 | 0.0502 |
| FREQ1 | 4.0390 | 0.1052 | 2181 | 0.0672 | 3.1369 | 1.5669 | 0.0261 |
| LITER3 | 0.4970 | 0.0160 | 2181 | 0.0107 | 0.5001 | 1.4915 | 0.0321 |
| NPYTDA | 0.3920 | 0.0133 | 2181 | 0.0105 | 0.4883 | 1.2763 | 0.0340 |

**Table 4.25: ANOVA on Means of Design Effects (*deft*) by Age Group and Variable Category**

| | | | Unique Method | | | | |
|---|---|---|---|---|---|---|---|
| | | | Sum of Squares | df | Mean Square | F | Sig. |
| DEFT | Main Effects | (Combined) | .198 | 6 | 3.303E-02 | 30.885 | .000 |
| | | AGEGRP | .124 | 2 | 6.218E-02 | 58.142 | .000 |
| | | CATEGORY | 7.381E-02 | 4 | 1.845E-02 | 17.256 | .001 |
| | Model | | .198 | 6 | 3.303E-02 | 30.885 | .000 |
| | Residual | | 8.555E-03 | 8 | 1.069E-03 | | |
| | Total | | .207 | 14 | 1.477E-02 | | |

**Table 4.26: ANOVA on Design Effects (*deft*) by Gender and Selected Variables, MIS**

| | | | Unique Method | | | | |
|---|---|---|---|---|---|---|---|
| | | | Sum of Squares | df | Mean Square | F | Sig. |
| DEFT | Main Effects | (Combined) | .800 | 13 | 6.157E-02 | 8.755 | .000 |
| | | AGEGRP | .223 | 2 | .112 | 15.886 | .000 |
| | | Selected Variables | .577 | 11 | 5.245E-02 | 7.458 | .000 |
| | Model | | .800 | 13 | 6.157E-02 | 8.755 | .000 |
| | Residual | | .155 | 22 | 7.032E-03 | | |
| | Total | | .955 | 35 | 2.729E-02 | | |

The ANOVA results show that at 0.05 level of significance, the *deft* values are significantly different among age groups (F=58.142, p = 0.000) and Variable Category (F=17.256, p = 0.001).

The analysis on selected variables further vindicates the differentials in the *deft* values across the age groups. The *deft* values are significantly different among the selected variables (F=7.458, p = 0.000).

## 4.7 Coefficients of Intra-Class Correlation (*roh*)

The computation of design effects for a survey enables one to compare the efficiency of the complex sample design adopted in relation to the use of simple random sampling. But, the design effects depend heavily on the size of sample clusters used and also the rate of homogeneity. The size of sample clusters may differ greatly for subclasses. The coefficient of intra-class correlation, *roh*, which measures the average degree of homogeneity within the PSUs of a particular variable is often used to model the sampling errors of complex designs. The coefficients of intra-class correlation (*roh*) by Variable Category and the selected variables in MIS are presented in Table 4.27 and Table 4.28 for total sample and geographical domains. The *roh* values by race, gender and age group subclasses are also presented in the tables.

The *roh* values for all 60 variables of total sample range from 0.05 to 0.24 and the average value is 0.12 for all variables under study. The Media Consumption category with an average *roh* value of 0.16 has the highest rate of homogeneity among the categories. Across geographical domains and subclasses, most of the average *roh* values in this category are around 0.2 except for RT stratum which is 0.11 and Malay subclass which is 0.15. The relatively large value of *roh* suggests that the media consumption habits do not vary much across the sample PSUs. However, the rate of homogeneity for FREQ1 is as high as 0.48 for the Chinese and 0.30 for Indian/Others subclass (see Table 4.28).

The lowest average rate of homogeneity for total sample is 0.08 for Household General category. Across geographical domains and subclasses, the highest average *roh* value in this category is the Indian/Others subclass (0.18), followed by Chinese subclass (0.14) and lastly MT stratum (0.13).

Apart from having large *roh* values in Media Consumption category, Chinese and Indian/Others subclasses also have high rate of homogeneity in other categories. For example, the average *roh* value for Durable Goods category in MT stratum is 0.15 and for Consumer Products in Malay subclass is 0.16. One of the factors contributing to the large value of *roh* for Durable Goods is the air conditioner ownership (DURA35) in MT stratum, which is highly correlated with type of dwelling units within the PSUs.

In the subclass design effect analysis, the *deft* values for Malays are higher than the Chinese and Indians/Others counterparts for all the Variable Categories especially for Consumer Products and Media Consumption categories. However, the *roh* value by ethnic group is greater for Indians/Others and Chinese compared to Malays for most of the categories and total sample. This illustrates that the design effects for Chinese and Indians/Others are higher than Malays if the same effective cluster size is used.

The average *roh* values for gender subclass do not vary much among Variable Category. Similar pattern is also observed for the average *deft* values in gender subclass. However, at the individual variable level, the difference between Males and Females in the *roh* values is more noticeable, especially for variables like NPYTDA, BANK13 and FREQ1.

In terms of age group subclass, the average *roh* values by Variable Category seems to be more even than average *deft* values because the average *roh* values for Durable Goods, Consumer Products and Household General categories are similar within category when compared across the 3 age groups. This is more reflective of the actual scenario as the variation in population distribution by age group should

not be too obvious until it causes significant difference on the above 3 categories that measure statistics on household.

For the remaining 2 categories that measure statistics based on individual respondents, the *roh* values are high (about 0.2) for all the age groups in Media Consumption. For the Individual General category, the *roh* values differ appreciably among the age groups especially for variables BANK13 and LITER3. For variable LITER3, the *roh* value for age group 15-29 is 0.06, whereas for age group 30-49 it is 0.17. But, the highest is among those aged 50 years and over, with *roh* value of 0.28. This is because those literate in Bahasa Malaysia who were born before independence in 1957 are only limited to specific group of population, resulting in high rate of homogeneity within sample PSUs.

Overall, the average *roh* values by geographical domain and subclass are below 0.20 with *roh* values for Chinese and Indian/Others subclasses just slightly below this value. The coefficients of intra-class correlation (*roh*) obtained from MIS fall within the expected range of *roh* values of most of the sample surveys (Kish et al., 1976).

**Table 4.27: Means Coefficients of Intra-Class Correlation (*roh*) of Variable Category by Geographical Domain and Subclass, MIS**

| Variable | Total Sample | Total MT | Total UT | Total RT | Total Malays | Total Chinese | Total Ind/Oth | Total Male | Total Female | Total 15-29 | Total 30-49 | Total 50+ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Durable Goods | 0.11 | 0.15 | 0.10 | 0.09 | 0.12 | 0.18 | 0.18 | 0.12 | 0.13 | 0.14 | 0.14 | 0.13 |
| Consumer Products | 0.12 | 0.11 | 0.13 | 0.12 | 0.16 | 0.17 | 0.18 | 0.14 | 0.14 | 0.15 | 0.16 | 0.13 |
| Household General | 0.08 | 0.13 | 0.08 | 0.05 | 0.08 | 0.14 | 0.18 | 0.11 | 0.10 | 0.10 | 0.10 | 0.07 |
| Media Consumption | 0.16 | 0.20 | 0.18 | 0.11 | 0.15 | 0.22 | 0.19 | 0.20 | 0.17 | 0.20 | 0.20 | 0.19 |
| Individual General | 0.10 | 0.09 | 0.11 | 0.10 | 0.13 | 0.16 | 0.11 | 0.12 | 0.11 | 0.11 | 0.15 | 0.15 |
| **Total Sample** | **0.12** | **0.14** | **0.12** | **0.10** | **0.13** | **0.18** | **0.17** | **0.14** | **0.13** | **0.14** | **0.15** | **0.14** |

**Table 4.28: Coefficients of Intra-Class Correlation (*roh*) of Selected Variables by Geographical Domain and Subclass, MIS**

| Variable | Total Sample | Total MT | Total UT | Total RT | Total Malays | Total Chinese | Total Ind/Oth | Total Male | Total Female | Total 15-29 | Total 30-49 | Total 50+ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DURA12 | 0.11 | 0.13 | 0.09 | 0.13 | 0.15 | 0.03 | 0.22 | 0.09 | 0.12 | 0.14 | 0.15 | 0.14 |
| DURA29 | 0.08 | 0.09 | 0.06 | 0.10 | 0.11 | 0.13 | 0.16 | 0.10 | 0.11 | 0.13 | 0.07 | 0.16 |
| DURA35 | 0.18 | 0.26 | 0.13 | 0.09 | 0.12 | 0.21 | 0.26 | 0.23 | 0.18 | 0.18 | 0.23 | 0.19 |
| P1MBUY15 | 0.15 | 0.14 | 0.14 | 0.18 | 0.23 | 0.20 | 0.31 | 0.20 | 0.19 | 0.18 | 0.19 | 0.23 |
| P1MBUY18 | 0.05 | 0.05 | 0.05 | 0.06 | 0.07 | 0.03 | 0.02 | 0.09 | 0.08 | 0.09 | 0.07 | 0.02 |
| MAABV15 | 0.07 | 0.09 | 0.06 | 0.05 | 0.07 | 0.18 | 0.12 | 0.10 | 0.08 | 0.10 | 0.07 | 0.05 |
| TVSET | 0.08 | 0.12 | 0.07 | 0.04 | 0.08 | 0.15 | 0.16 | 0.10 | 0.09 | 0.10 | 0.10 | 0.11 |
| FREQ1 | 0.24 | 0.31 | 0.30 | 0.16 | 0.14 | 0.48 | 0.30 | 0.30 | 0.26 | 0.31 | 0.27 | 0.34 |
| NPYTDA | 0.12 | 0.15 | 0.17 | 0.09 | 0.16 | 0.09 | 0.24 | 0.10 | 0.15 | 0.15 | 0.15 | 0.15 |
| ACTVT1 | 0.08 | 0.10 | 0.07 | 0.05 | 0.09 | 0.17 | 0.06 | 0.10 | 0.08 | 0.10 | 0.13 | 0.15 |
| BANK13 | 0.10 | 0.09 | 0.13 | 0.09 | 0.14 | 0.14 | 0.06 | 0.15 | 0.10 | 0.12 | 0.16 | 0.09 |
| LITER3 | 0.11 | 0.10 | 0.13 | 0.11 | 0.09 | 0.16 | 0.04 | 0.12 | 0.15 | 0.06 | 0.17 | 0.28 |

**Table 4.29: ANOVA on Means of Coefficients of Intra-class Correlation (*roh*) by Geographical Domain and Variable Category**

| | | | Unique Method | | | | |
|---|---|---|---|---|---|---|---|
| | | | Sum of Squares | df | Mean Square | F | Sig. |
| ROH | Main Effects | (Combined) | 1.372E-02 | 6 | 2.287E-03 | 3.191 | .066 |
| | | CATEGORY | 9.218E-03 | 4 | 2.304E-03 | 3.215 | .075 |
| | | STRATUM | 4.506E-03 | 2 | 2.253E-03 | 3.143 | .098 |
| | Model | | 1.372E-02 | 6 | 2.287E-03 | 3.191 | .066 |
| | Residual | | 5.735E-03 | 8 | 7.169E-04 | | |
| | Total | | 1.946E-02 | 14 | 1.390E-03 | | |

**Table 4.30: ANOVA on Coefficients of Intra-class Correlation (*roh*) by Geographical Domain and Selected Variables, MIS**

| | | | Unique Method | | | | |
|---|---|---|---|---|---|---|---|
| | | | Sum of Squares | df | Mean Square | F | Sig. |
| ROH | Main Effects | (Combined) | .113 | 13 | 8.702E-03 | 5.676 | .000 |
| | | Selected Variables | .104 | 11 | 9.411E-03 | 6.139 | .000 |
| | | STRATUM | 9.606E-03 | 2 | 4.803E-03 | 3.133 | .064 |
| | Model | | .113 | 13 | 8.702E-03 | 5.676 | .000 |
| | Residual | | 3.373E-02 | 22 | 1.533E-03 | | |
| | Total | | .147 | 35 | 4.196E-03 | | |

A 2-way analysis of variance (ANOVA) is performed to evaluate the observed differences of *roh* values in Table 4.27 and Table 4.28.

Results show that at 0.05 level of significance, the average *roh* values are not significantly different among the Variable Category ($F=3.215$, $p = 0.075$). The geographical domain is also not a significant factor in determining the average *roh* value ($F=3.143$, $p = 0.098$). This suggests that the rate of homogeneity of sample PSUs is equal in all strata.

**Table 4.31: ANOVA on Means of Coefficients of Intra-class Correlation (*roh*) by Ethnic Group and Variable Category**

| | | | Unique Method | | | | |
|---|---|---|---|---|---|---|---|
| | | | Sum of Squares | df | Mean Square | F | Sig. |
| ROH | Main Effects | (Combined) | 1.228E-02 | 6 | 2.046E-03 | 2.941 | .080 |
| | | CATEGORY | 6.528E-03 | 4 | 1.632E-03 | 2.346 | .142 |
| | | RACE | 5.749E-03 | 2 | 2.874E-03 | 4.132 | .059 |
| | Model | | 1.228E-02 | 6 | 2.046E-03 | 2.941 | .080 |
| | Residual | | 5.566E-03 | 8 | 6.957E-04 | | |
| | Total | | 1.784E-02 | 14 | 1.274E-03 | | |

**Table 4.32: ANOVA on Coefficients of Intra-class Correlation (*roh*) by Ethnic Group and Selected Variables, MIS**

| | | | Unique Method | | | | |
|---|---|---|---|---|---|---|---|
| | | | Sum of Squares | df | Mean Square | F | Sig |
| ROH | Main Effects | (Combined) | .183 | 13 | 1.405E-02 | 2.572 | 025 |
| | | RACE | 1.447E-02 | 2 | 7.233E-03 | 1.324 | 286 |
| | | Selected Variables | .168 | 11 | 1.529E-02 | 2.799 | 019 |
| | Model | | .183 | 13 | 1.405E-02 | 2.572 | 025 |
| | Residual | | .120 | 22 | 5.464E-03 | | |
| | Total | | .303 | 35 | 8.654E-03 | | |

The analysis on the effect of ethnic group and Variable Category (Table 4.31) on the *roh* values show that at 0.05 level of significance, the average *roh* values among ethnic groups are not significantly different (F=4.132, p = 0.059) The Variable Category is also not a significant factor in determining the average *roh* values (F=2.346, p = 0.142). The result in Table 4.32 gives similar conclusion where ethnic group is not a significant factor in determining the *roh* values (F=1.324, p = 0.286) but, the *roh* values are significantly different among the selected variables (F= 2.799, 0.019).

**Table 4.33: ANOVA on Means of Coefficients of Intra-class Correlation (*roh*) by Gender and Variable Category**

| | | | Unique Method | | | | |
|---|---|---|---|---|---|---|---|
| | | | Sum of Squares | df | Mean Square | F | Sig. |
| ROH | Main Effects | (Combined) | 7.867E-03 | 5 | 1.573E-03 | 20.694 | .006 |
| | | CATEGORY | 7.633E-03 | 4 | 1.908E-03 | 25.098 | .004 |
| | | SEX | 2.343E-04 | 1 | 2.343E-04 | 3.081 | .154 |
| | Model | | 7.867E-03 | 5 | 1.573E-03 | 20.694 | .006 |
| | Residual | | 3.041E-04 | 4 | 7.603E-05 | | |
| | Total | | 8.171E-03 | 9 | 9.079E-04 | | |

**Table 4.34: ANOVA on Coefficients of Intra-class Correlation (*roh*) by Gender and Selected Variables, MIS**

| | | | Unique Method | | | | |
|---|---|---|---|---|---|---|---|
| | | | Sum of Squares | df | Mean Square | F | Sig. |
| ROH | Main Effects | (Combined) | 7.965E-02 | 12 | 6.637E-03 | 12.781 | .000 |
| | | Selected Variables | 7.931E-02 | 11 | 7.210E-03 | 13.884 | .000 |
| | | SEX | 3.375E-04 | 1 | 3.375E-04 | .650 | .437 |
| | Model | | 7.965E-02 | 12 | 6.637E-03 | 12.781 | .000 |
| | Residual | | 5.713E-03 | 11 | 5.193E-04 | | |
| | Total | | 8.536E-02 | 23 | 3.711E-03 | | |

The ANOVA results show that at 0.05 level of significance, gender is not a significant factor in determining the *roh* values (Table 4.33: F=3.018, p = 0.154; Table 4.34: F=0.650, p = 0.437). However, the average *roh* values among the Variable Category are significantly different (F=25.098, p = 0.004) as well as among selected variables (F=13.884, p = 0.000). Above results strongly suggest that the rate of homogeneity between Males and Females is not significantly different.

**Table 4.35: ANOVA on Means of Coefficients of Intra-class Correlation (*roh*) by Age Group and Variable Category**

| | | | Unique Method | | | | |
|---|---|---|---|---|---|---|---|
| | | | Sum of Squares | df | Mean Square | F | Sig |
| ROH | Main Effects | (Combined) | 1.751E-02 | 6 | 2.918E-03 | 12.908 | 001 |
| | | AGEGRP | 5.143E-04 | 2 | 2.571E-04 | 1.138 | 367 |
| | | CATEGORY | 1.699E-02 | 4 | 4.248E-03 | 18.793 | 000 |
| | Model | | 1.751E-02 | 6 | 2.918E-03 | 12.908 | 001 |
| | Residual | | 1.808E-03 | 8 | 2.260E-04 | | |
| | Total | | 1.931E-02 | 14 | 1.380E-03 | | |

**Table 4.36: ANOVA on Coefficients of Intra-class Correlation (*roh*) by Age Group and Selected Variables, MIS**

| | | | Unique Method | | | | |
|---|---|---|---|---|---|---|---|
| | | | Sum of Squares | df | Mean Square | F | Sig |
| ROH | Main Effects | (Combined) | .147 | 13 | 1.134E-02 | 6.436 | 000 |
| | | AGEGRP | 2.639E-03 | 2 | 1.319E-03 | .749 | 485 |
| | | Selected Variables | .145 | 11 | 1.316E-02 | 7.470 | 000 |
| | Model | | .147 | 13 | 1.134E-02 | 6.436 | 000 |
| | Residual | | 3.876E-02 | 22 | 1.762E-03 | | |
| | Total | | .186 | 35 | 5.319E-03 | | |

The analysis on age group gives similar conclusion as gender where the *roh* values are not significantly different among age groups. There is no evidence to conclude that the population is unevenly distributed in terms of age groups across the sample PSUs in MIS. However, the average *roh* values among the Variable Category are significantly different ($F=18.793$, $p = 0.000$) as well as among selected variables ($F=7.470$, $p = 0.000$).

In summary, the analyses on *roh* values show that the rate of homogeneity of PSUs is primarily depends on the variables for the survey. Geographical domains and ethnic groups may have little effects on the rate of homogeneity. However, gender and age groups are not the significant factors in determining the rate of homogeneity.