**Survey data analysis**

**Task1. Survey Info**

Survey date: 03/06 – 06/06

Survey site: *WeChat*

Sampling method: Snowball

Reached: 1587 (7 provincial level district in China)

Returned: 247 (189 participants are from Anhui Province where the survey started to circulate in WeChat networks)

Effective sample: 225

Mean survey completion time: 14mins, 27seconds

**Task2. Data cleaning - DONE**

**Task3: Data description**

*In the very beginning, I am providing basic descriptive statistics for the data.*

# mothers who are only-children FOR their own parents **[only\_child]**

*table(dat$only\_child)*

*FALSE TRUE*

*146 79*

# mothers who are young mothers (age <= 36) **[young\_mother]**

table(dat$young\_mother)

FALSE TRUE

77 148

# mothers who have only one child **[one\_child]**

table(dat$one\_child)

FALSE TRUE

69 156

=== because there are more young mothers

# mothers who want to have a 2nd child **[self\_2\_child]**

table(dat$self\_2\_child)

FALSE TRUE

145 80

# mothers’ education levels **[education]**

1 = high school or lower; 2 = 3-year college; 3 = university or higher

table(dat$education)

1 2 3

72 83 70

# household income level – average monthly income per capita **[income]**

1 = <6000; 2 = 6000 ~ 10000; 3 = > 10000

table(dat$income)

1 2 3

90 63 72

# mothers’ status in original family where they grew up **[status\_origin]**

summary(dat$status\_origin)

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.000 2.857 3.571 3.514 4.571 5.000

> boxplot(dat$status\_origin, main = "mothers status in original family")



# mothers’ status in current family including the in-laws **[status\_current]**

summary(dat$status\_current)

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.000 3.000 4.000 3.708 5.000 5.000

> boxplot(dat$status\_current, main = "mothers status in current family")

**On average mothers claim higher status in current family (mean = 3.71) than original family (mean = 3.51). (one-child policy women’s liberation) –**

**=== think about the correlation with education**

# mothers’ childhood experience/feelings for only-children

summary**(dat$childhood)**

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.000 2.667 3.333 3.351 4.667 5.000

> boxplot(dat$childhood, main = "mothers childhood feelings for only-children")

**On average, mothers agree that being-only-children has better education (mean = 3.51)/living condition (mean = 3.61); however, they do not think being-only-children are happier nor less happy (mean = 2.82 for ‘do you think being only-children is happier’).**



=== only children: disjunction between material and human side; they are missing on the social side, but gaining on other. (whether the this finding vary between other variables)

# life satisfaction

summary(**dat$life\_satisfaction)**

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.000 2.333 3.000 3.259 4.000 5.000

> boxplot(dat$life\_satisfaction, main = "mothers' life satisfaction")

for subscales do reliability testing – done in R



# motherhood satisfaction

summary(dat$motherhood\_satisfaction)

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.000 2.875 3.375 3.385 4.000 5.000

> boxplot(dat$motherhood\_satisfaction, main = "mothers' motherhood satisfaction")

On average, the mean score is 3.39 (average over 8 items). Mothers have more or less satisfactory motherhood experience. However, satisfaction on item 4 = ‘I’m doing better than average mother’, 6 = ‘husband is supportive’, 8 = ‘parents-in-law are supportive’ are relatively lower. (3.17, 3.19 and 2.81 respectively). Interestingly, mothers claim better status in current family. However, they do not appreciate support from in-law family as much as original family. (contradict to Chinese tradition – in-law family has far more responsibility for offspring; but more likely due to that mothers didn’t feel being supported when they are actually supported. – a kind of discounting in-law family or **miss-interpretation of in-laws)**.你妈给你十块钱和你婆婆给你十块钱一样吗？你妈说你滚和你婆婆说你滚是一样吗？这里可以聊下culutre。

=== If in-laws is significant across samples, there might be a shift in terms of current culture of mothering/parenting.

===

# motherhood interferes work

summary**(dat$mother\_interferes\_work)**

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.000 1.000 2.000 **2.314** 3.000 5.000

> boxplot(dat$mother\_interferes\_work, main = "thinking motherhood interferes work")



# work interferes motherhood

summary**(dat$work\_intergeres\_mother)**

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.000 1.000 2.333 **2.487** 3.000 5.000

> boxplot(dat$work\_intergeres\_mother, main = "thinking work interferes motherhood")

/// !! typo in naming scheme – interferes, not intergeres!!

/// On average, mothers tend to believe that work and motherhood do not interfere each other. However, more concern is given to motherhood, rather than work. **Compare mothers’ ranking of self/mother/wife/daughter/employee, we can see in average, mothers rank ‘being a mother’ the most important, and ‘being an employee’ the least important. – (母亲第一，工作最后: this is consistent with traditional culture, which define women’s main responsibilities of childrearing/childbearing and child education)---虽然大部分中国女性接受了高等教育，但文化的影响却是根生蒂固。**

[1 = most important; 5 = least important]

> mean**(dat$rank\_occup)**

[1] 4.288889

> mean(**dat$rank\_mother)**

[1] 2.017778

> mean**(dat$rank\_self)**

[1] 2.773333

> mean**(dat$rank\_daughter)**

[1] 2.817778

> mean**(dat$rank\_wife)**

[1] 3.102222

**mother > self > daughter > wife > occupation**

**=== because they have children, so they thing along that way. If the same women have done the interview before they had kids, we might come out with less conventional outcomes than here.**

**=== clustering – mother, occupation and other**

**two sample t test (alpha = 0.05)**

**next – expand to how different between groups**

**- descriptive – group mean difference**

**- then maybe linear or logistic model**

# group by whether mothers are only-children to their own parents (是独生子女和非独生子女两个group与以下变量的相关性dependent)

I find group differences in the following variables

group <- "only\_child"

unlist(lapply(names(dat)[-c(1:3)], function(i) two\_sample\_t(i, group, dat)))

[1] "fertility\_preference" "self\_2\_child" "motherhood3"

[4] "rank\_daughter" "mother\_birth\_order" "education"

[7] "one\_child" "young\_mother" "childhood"

**"fertility\_preference" – use 1-5 rate “I prefer having only one child.”**

**“self\_2\_child” – the mother wants to have a 2nd child Yes/No**

**"rank\_daughter" - rank the importance of being a daughter to the mother using 1-5**

**"childhood" – childhood feelings over “being only children have better living condition / education and are happier”**

# **group by whether mothers have only one child**

# we find group difference in the following variables

group <- "one\_child"

unlist(lapply(names(dat)[-c(1:3)], function(i) two\_sample\_t(i, group, dat)))

[1] "fertility\_preference" "self\_2\_child" "husband\_2\_child"

[4] "parents\_2\_child" "child\_2\_child" "only\_child"

[7] "childhood" "pro\_2nd\_child"

# **group by whether mothers are willing to have a 2nd child**

# we find group difference in the following variables

group <- "self\_2\_child"

unlist(lapply(names(dat)[-c(1:3)], function(i) two\_sample\_t(i, group, dat)))

[1] "fertility\_preference" "husband\_2\_child" "child\_2\_child"

[4] "rank\_wife" "only\_child" "one\_child"

[7] "young\_mother" "childhood" "con\_2nd\_child"

# **group by whether mothers are young (<= 36 years )**

# we find group difference in the following variables

group <- "young\_mother"

unlist(lapply(names(dat)[-c(1:3)], function(i) two\_sample\_t(i, group, dat)))

[1] "self\_2\_child" "parents\_2\_child" "inlaws\_2\_child"

[4] "rank\_self" "rank\_daughter" "mother\_birth\_order"

[7] "education" "only\_child"

**### chi-square test (I need to do some more.)**

**== do logistic**

Make a two-way table of ‘education’ by ‘whether a mother is herself an only-child’. And do a Chi-squared test of independence.

mytable <- table( young = dat$young\_mother,

education = dat$education)

summary(mytable)

Number of cases in table: 225

Number of factors: 2

Test for independence of all factors:

Chisq = 23.212, df = 2, p-value **= 9.111e-06**

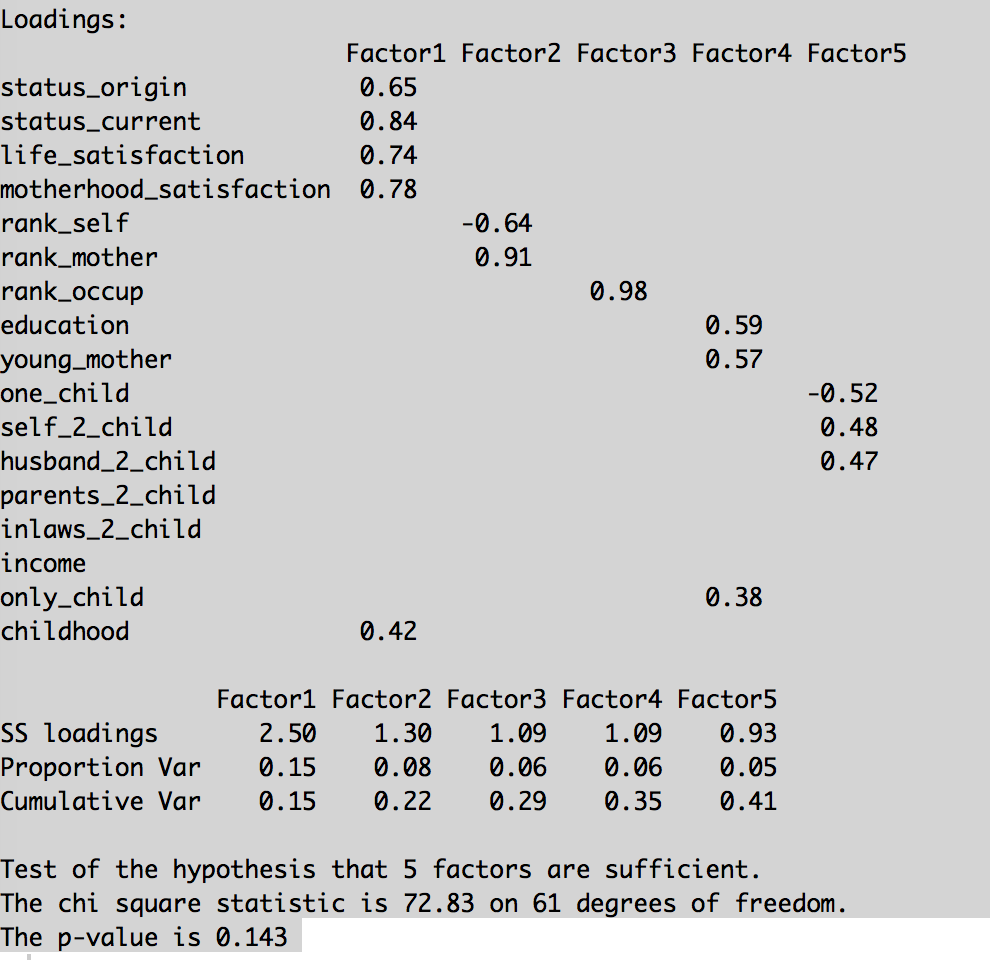
‘**education’ by ‘whether a mother is herself an only-child’ are dependent.**

**# exploratory factor analysis**

# I am not yet clear how to do it. There are some R codes shared online. So I just borrowed and used.

d.fac <- factanal(d2, factors = 5, rotation = "varimax", scores=c("regression"))

print(d.fac, digits = 2, cutoff = .3, sort = TRUE)



How I can understand here is:

Using 5 factors (each cluster a few inter-correlated variables) can fit the data good enough. What is next? Cluster relevant variables for each factor, and use average score to represent each factor, and then do such as regression modelling?

==

== content analysis

in an ideal world, identify 3 or 4 categories: traditional, mixed, modern motherhood. Then do logistic regression using these as predict variable for the other variables.

Local cultural variables