Analysis of contraceptive method used

Abstract: Contraceptive methods are important in control of population. In this research, logistic regression and multinomial logit model are used to evaluate how the women used contraceptive method. It is found that younger women with low education background and Islamic women has higher chance to prefers no use of methods.

1. Introduction

1.1 Population explosion

Population explosion means that the poplution in the world has rapidly increased. For example, the population in China has increased to 1.42 billions in 2020 when only 0.8 bilions in the 1970s. As we all known that there is limited resource in the earth, with population increased, the average resource shared per person will decrease, such as food.

1.2 Population Planning

Population Planning is essential. The contraceptive method is one of the ways to limit the increase in population. Therefore, analysis of contraceptive method is very important for one government desired to proceed population planning. For example, if low-income family has no knowledge on the contraceptive method, the number of children they owned may be led to different from the high-income family.

2. Data Description

There are 1473 obsevations with no missing value in the data set. 629 (42.7%) interviewees responded that they did not use contraceptive method when 333 (22.6%) and 511 (34.69%) interviewes used short-term methods and long-term methods respectively. The variables in dataset are wife's age (16 -49, integer), wife's education (1-4 by ascending order and 1 is the lowest and 4 is the highest), husband's eduation, number of children born, wife's religion which 0 represented non-islam and 1 is for islam. Binary variable to show work status for the women and her husband's occupation (1,2,3,4) were asked in the interview. Living standard index (1,2,3,4) and media exposure (binary with 0 for good and 1 for not good) are also in the data set.

Wife's age and number of children born are only numerical variables.
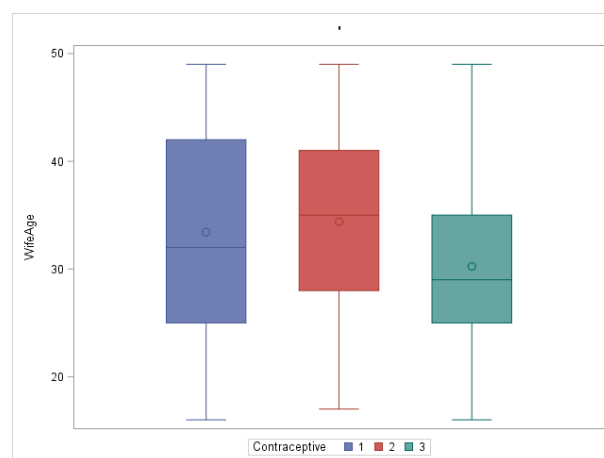

Fig 1. The box plot above showed the relationship between wife's age and contraceptive methods.

In figure 1, it demonstrated the rough distribution for wife with different contraceptive methods. It showed that in the pool of adopting short-term, the average of wife's age is apparently lower than the other two groups. It may lead a picture or motivation to investigate why the younger wife may choose short-term contraceptive methods.
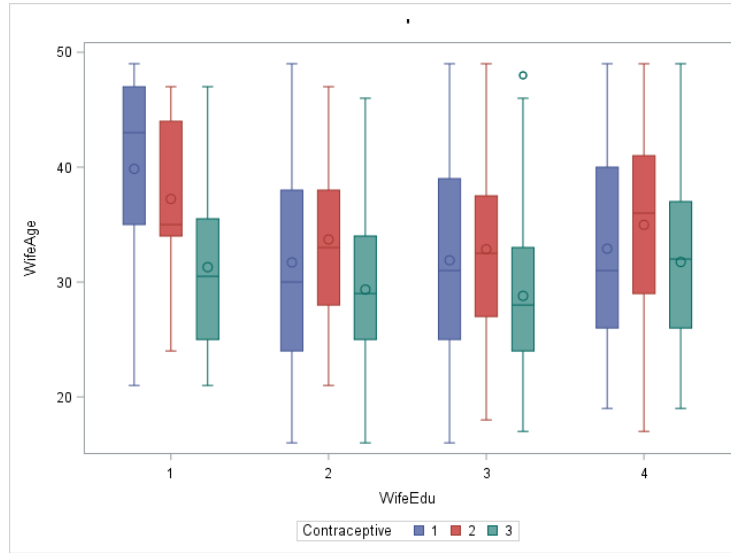
Fig 2..Box plot of wife's age with different eduation background

In figure 2, it is noticeable that in low-level education background, women who adopted short-term method is obviously are younger than the other women with two methods. This phenomenon is interesting.

3. Method

3.1 Logistic regression

Logistic regression is used for binary response. Even though the response (contraceptive method) is not binary, the response can be re-organised to be binary. Here, the research should be investigated how the other variables affect how they used contraceptive method. Therefore, we can merge the response from {long-term method, short-term method} to only {used contraceptive method}, i.e. set 1 if they used contraceptive method and set 0 if they did not do so. By logit function,

$$log(\frac{P(Y = 1|X)}{1 - P(Y = 1|X)}) = X\beta$$

Where Y follow Bernoulli dribution with probability p, i.e. Y ~ Bern(p). X is the design matrix for prediction and β is regression coefficient.

3.2 Multimonial logit regression

Regression can be extended by logistical regression. The link function ($g(\mu)$) will be changed. Before that, reference group is taken by J^{th} group. For example, we can take long-term method as reference.

Odds in mulitmonial logit regression is defined as:

$$\frac{I_j(x)}{I_J(x)} = X\beta$$

Where $j = 1…2…J-1.$ and $I_j(x)$ = P(Y = j|X). and Y ~ Multinomial(p_1, p_2, p_3)

The regression model is denoted by $log(\frac{I_j(x)}{I_J(x)}) = X\beta$

4. Results

4.1 Analysis of variables and the contraceptive methods

By hypothesis,

$$\begin{cases} H_0: Working\ status\ is\ insignificant\ to\ the\ woman\ whether\ adopting\ conceptive\ method. \\ H_1: Working\ status\ is\ significant \end{cases}$$

It can be transformed to our logistic model.

In logistic model, X is only including $I_{(working\ status\ =\ 0)}$ and β only includes constant term and β_1

Such that $g(\mu) = \beta_0 + \beta_1 * I_{(working\ status\ =\ 1)}$

Therefore, the Hypothesis changed to

$$\begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{cases}$$

As if β_1 = 0, the probability of adopoting conceptive methive is estimated to be constant which is not related to working status.

R-Square	0.0015	Max-rescaled R-Square	0.0021

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	2.2745	1	0.1315
Score	2.2832	1	0.1308
Wald	2.2804	1	0.1310

By performing Likelihood ratio test, p-value is 0.1315 > $\alpha = 0.05$. We do not reject $\beta_1 = 0$. Therefore, we do not reject that Working status is insignificant to woman whether adopting conceptive method at $\alpha = 0.05$.

By the similar hypothesis, but the target changed to Islamic women.

R-Square	0.0054	Max-rescaled R-Square	0.0073

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	7.9881	1	0.0047
Score	7.8379	1	0.0051
Wald	7.7604	1	0.0053

By performing Likelihood ratio test, p-value is 0.0047 > $\alpha = 0.05$. We do not reject $\beta_1 = 0$. Therefore, we do not reject that Working status is insignificant to woman whether adopting conceptive method at $\alpha = 0.05$

By the simple model above, it is suspected that the religious effect may larger than working status effect. Therefore, different variables contains different effect on the contraceptive method.

By this approach, it can be found the significant variables. As the amounts of categorical variables are 7 and hard to fit the full model with interaction terms. Therefore, this approach is equalvalent to forward selection which stops at no significant variables.

4.2 Logistic regression model

Final variables set = {`WifeEdu numChild WifeAge WifeAge*numChild numChild*WifeEdu LivingStandard numChild *LivingStandard Religion`}

However, p-value of deviance and Pearson Goodness-of-Fit statistics are <0.0001 and 0.0005 respectively. Therefore, the model is considered as invalid.

We consider transformation in the model to enable statistics to be fulfilled, which is Log-transformation on number of Children born + 1 and wife's age.

New variables set = {`WifeEdu, log(numChild+1), log(WifeAge), log(WifeAge)* log(numChild+1), log(numChild+1)*WifeEdu, LivingStandard, Religion, MeidaExp`}
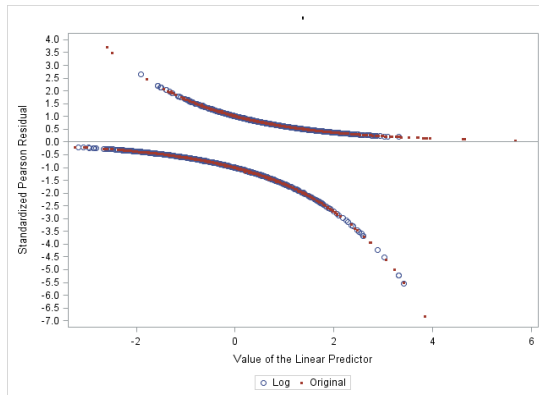
p-value of deviance and Pearson Goodness-of-Fit statistics are 0.0003 and 0.0260 respectively. Therefore, the model is considered as invalid but the p-value is higher than before.

By the residual plot of two models, it was found that Log-transformed model has fewer suspected outliers. It maybe the reason why p-value is higher.

The amounts of outliers in log-transformed logistic model is 14. After deleting outliers and refit the model. The model was changed. New variables set = {`WifeEdu,log(numChild+1)` `log(WifeAge),log(WifeAge)*log(numChild+1),log(numChild+1)*WifeEdu,` `LivingStandard,Religion, MeidaExp, log(WifeAge)*WifeEdu`}. P-value of Peason Chi-squared statistic is 0.4381 which means we cannot reject the linear relationship between predictors and logit function. Regarding the influential points, all suspected influential point cannot make significant changes in coefficients by DFBETA.
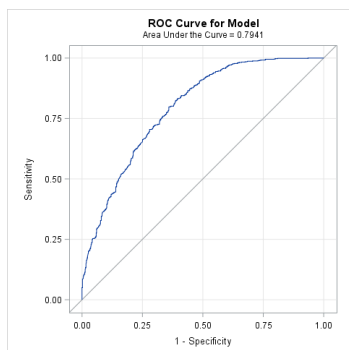
To evaluate the model, Area under ROC curve is ≈ **0.8** which indicated that the model is considered as the moderate performed.

It provided confidence to use the model to evaluate or explain phenomenon.

4.3 Multinomial logit model

By logistic model, it showed that only interaction term between two factors are significant. Theorefore, we simplify that there are no higher order interaction terms more than 2.

After fitting the full model, p-value of deviance and Pearson Goodness-of-Fit statistics are 0.9999 and 0.0781. After forward selected model, p-value of deviance and Pearson Goodness-of-Fit statistics are 0.9904 and 0.2638. Therefore, both models are valid. Again, log-transformed model was performed. P-value of deviance and Pearson Goodness-of-Fit statistics are 0.9997 and 0.8253.

To compare two models, AIC (Akaike Information Criterion) and SC (Schwarz Criterion) are used.

Criterion	Original model	Log-transformed
AIC	2794.709	2717.651
SC	3027.691	2950.63

By the table on the left, log-transformed model is better in our training data set. Therefore, the following evaluation would be using the log-tranformed model as the logistic regression did before.

Variable set = { `log(WifeAge), log(numChild+1), Religion, LivingStandard,` `Husband's occupation, Wife's Education, Husband's Education,log(WifeAge)*` `Husband's Education,log(numChild+1)* Wife's Education` }

Joint Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
log_WifeAge	2	13.6998	0.0011
log_numChild	2	12.0321	0.0024
WifeReli	2	8.5945	0.0136
LivingIndex	6	17.3552	0.0081
HusOccu	6	13.4433	0.0365
HusEdu	6	13.5824	0.0347
WifeEdu	6	5.7787	0.4484
log_WifeAge*HusEdu	6	12.9495	0.0438
log_numChild*WifeEdu	6	12.7597	0.0470

Coefficient table for multinomial logit Model.

4.4 Hypothesis testing

4.4.1 Analysis on wife's education and log(age)

From now on, the data will be interpreted by the hypothesis testing.

According to Figure 2, it cannot easily show that higher education increases the probability of adopting contraceptive method. The following statement is what we wanted:

The odds of Ladies with higher education to choose contraceptive method relative to no method used is higher.

By logistic model, the table showed the odds ratio below.

| Education (at log(child + 1) = 1.2903 and log(age) = 3.4508	Estimate	95% Wald C.I.
2 vs 1(reference)	1.004	[0.593, 1.701]
3 vs 1(reference)	1.507	[0.895, 2.536]
4 vs 1(reference)	5.176	[2.985, 8.974]

According to odds ratio, the odds of education = 2 or edication = 3 relative to the wife with education = 1 chooses contraceptive method cannot be rejected that they are equal as 1.0 is in the confidence interval. However, probability of wife with education = 4 relative to the education = 1 is estimated higher 417.6% and 1 is not in 95% confidence interval so that we can reject the probabilities is equal at mean of log(child + 1) and log(age).

However, this phenomenon is not valid for every cases as log(child + 1) and log(age) are another two factor to determine the odds ratio as they are interact with education in model.

By hand-on calculation, odds ratio given log(child +1) and log (age) when education =2 vs 1

= $e^{-6.7402+1.9481*log(age)+0.0618\,log(child+1)}$. It showed that at higher age and more child born ever would increase the odds ratio so that at some points the estimate of odd ratio will increase
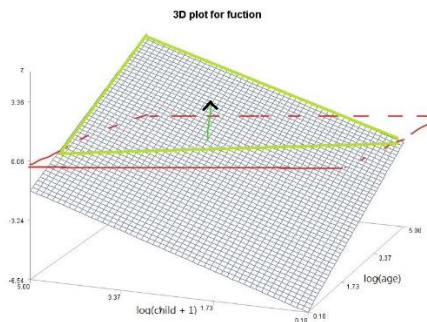
The yellow region showed the solution of the estimate > 1 which indicate the probability of education = 2 relative to education = 1 at some points on log(child +1) and log (age) chooses contraceptive methods is equal.

The statistical result is only obvious to showthe higher education level would have high chance to adopt conceptive method relative to low education level at some points of child and age but not the general trend.

In the following paragraph, the multinomial logit regression was done for analyzing log(age).

According to Figure 2, it is observed that if the wife's education is low, there are more chances that the young lady would choose short term contraceptive method.

It is equalvant to that the probability of younger lady with low education background choosing short term contraceptive method would higher older lady with low education. Then we can do hypothesis testing with multinomial logit model. (relative to Y = 1)

$$\begin{cases} H_0: Odds\ that\ younger\ ladies\ with\ low\ education\ choose\ Y = 3 > \\ Odds\ that\ older\ ladies\ with\ low\ education\ choose\ Y = 3 \\ H_1: Otherwise \end{cases}$$

By odd ratios of log(age) + 1 at low education with log(age) at low education choose contraceptive method =3 relatively to method =1 is

$$= \begin{cases} e^{-7.1966 + 1.4873} = 0.00331, if\ her\ husband\ edcuation\ level = 2 \\ e^{-7.1966+ 2.9085} = 0.0137, if\ her\ husband\ education\ level = 3 \\ e^{-7.1966 + 3.6326} = 0.0283, if\ her\ husband\ education\ level = 4 \end{cases}$$

As there is no interaction term between between log(age) and wife's education in multinomial logit model, the null hypothesis becomes Odds that younger ladies choose Y=3 relative to Y = 1 > Odds that older ladies choose Y=3 relative to Y = 1 and it just depends on husband education level in our model. Therefore, it is expected probability that the respondent with age 31 comparing to age 30 chooses short-term method relative to no contraceptive method used would be decrease

$$=$$
$$(1 - e^{-7.1966*ln(31/30) + 3.6326*ln(31/30)}) = 11.029\%$$ if her hushand education level = 3. It provides the rough idea of probability of younger wife would really chooses short-term methods preferably. It may be consistent with our observation from figure 2

Odds Ratio Estimates and Wald Confidence Intervals		
Odds Ratio	Estimate	95% Confidence Limits
Contraceptive 3: log_WifeAge at HusEdu=1	<0.001	<0.001	0.186
Contraceptive 3: log_WifeAge at HusEdu=2	0.003	<0.001	0.016
Contraceptive 3: log_WifeAge at HusEdu=3	0.014	0.004	0.043
Contraceptive 3: log_WifeAge at HusEdu=4	0.028	0.012	0.068

Table: Odds ratio estimate and its wald confidence intervals

Not only estimates, there are also the 95% confidence limits shows that there is lower chance of the older respondents to choose short term method relative to no method used.

4.4.2 Analysis of religion

Religion is well-known that it affected the sexual activity. For example, Catholicismus does not allow sex before marriage. Therefore, it is exiting whether it affect how women used contraceptive methods.

By logistic model,

Religion	Estimate	95% Wald C.I.
Islamic v.s. non-Islamic	0.618	[0.425, 0.897]

The odds that the Islamic respondents relative to non-Islamic will decrease 38.2% so

$\frac{P(Y = 1)}{P(Y= 0)}$ will decrease 38.2% if she is Islamic. It showed that there is lower chance that she would like to adopt contraceptive method. In 95% Wald confidence interval, the odds will decrease [57.5%, 10.3%] relative to non-islamic woman. And also in the model, p-value of religion is 0.0097<0.05 so it is significant in the model.

By multinomial logit model,

Religion	Mehtod	Estimate	95% Wald C.I.
Islamic v.s. non-Islamic	2	0.536	[0.325, 0.818]
Islamic v.s. non-Islamic	3	0.666	[0.442,1.004]

$$\begin{cases} H_0: Odds\ ratio\ =\ 1 \\ \quad H_1: Otherwise \end{cases}$$

When the respondent is Islamic, the odds that the she has chosen long-term contraceptive method relative to no contraceptive method is higher in both expected and even confidence limit. It is expected to be decrease 46.4%. We reject H_0 because 1 is out of confidence limit. However, we do not reject H_0 when the odds ratio which is choosing short-term contraceptive method relative to no contraceptive method used is not significantly different from 1 but 1.004 Is very marginal limit.

Therefore, The multinomial regresision is almost consistent with the odds ratio in logistic regression.

5. Discussion

Overall, multinomial logit told us that younger with no matter husband's education background has higher probability to select short-term contractive method relative to no methods used. And logistical regression provided us the odds of the woman with higher education adopted contraceptive method relative to no methods used but the relationship depends on the log(age) and log(child +1). It is harder to suggest one conclusion without log(age) and log(child +1). Analysis of religion showed the Islamic woman really prefers no used comparative to non-islamic woman in logistic regression. Analyzing similar situation by multinomial logit regression, it provides clear picture to know odds and realize the relationship between long-term methods and no contraceptive method and the relationship between short-term methods and no contraceptive method.

However, it contains limitation.

Even though ROC curve of logistic regression is around 0.8 to provide confidence to use this model to evaluation, R^2 is only 0.2662 which is extremely low and p-vlaue of deviance goodness-of-fit test statistic is 0.02 <0.05, though p-value of Pearson's Chi-squared statistic is around 0.5.

Multinomial regression may not be consistent with the logistic model by observation. From the religion analysis, confidence interval included 1 when contraceptive method is 3 shown above. However, similar study of religion in logistic regression rejected the odds that Islamic woman has higher chances to select contraceptive methods respect to no contraceptive while multilogit regression does not concludes in both situations, i.e. Y = 2 and Y =3 cases.

By the analysis, it is suggested that we should educate the Islamic woman to use contraceptive methods. For example, the government and social institutions should provides more talks to the Islamic. Younger woman with low education standard prefers short-term method. However, short-term method may not be safer than long-term methods, such as the efficiency of condemn based on whether we know how to use. Long-term methods does not exist this problem if we remember to take pills.