Application of Censored Regression Model to Factors Responsible for Extra-Marital Affairs

Bello, A. H[1]. and Alabi, O.O.[2]

[1]Department of Statistics, School of Sciences, Federal University of Technology, Akure. Ondo –State. Nigeria.
[2]Department of Statistics, School of Sciences, Federal University of Technology, Akure. Ondo –State. Nigeria.
E-Mail: habello@futa.edu.ng

Abstract: The study is on the application of censored regression analysis to the factors responsible for extramarital affairs among married staff in the school of sciences and school of agriculture of Federal University of Technology Akure (FUTA). Questionnaire design technique was used to collect relevant data from the respondents. The methodology employed was censored regression to fit the model and Maximum Likelihood Estimator (MLE) was used to obtain the parameters of the estimate. The results from the study revealed that age in marriage, gender response and religion are the main factors that determine extramarital affairs. They are statistically significant with p-value < 0.05. Other factors such as age, education qualification and number of years married are not significant as they contribute less to extra marital affairs. The fitted model is Marital Affairs = 10.69902 – 0.521391 (Education) + 0.540626 (Religion) + 0.927884 (Rating) + 2.188614 (Gender) −0.028957(Age) −0.038507(Year Married). The computed R squared (R^2) value of 65% indicate that the model is moderately fit.
[Bello, A. H. and Alabi, O.O. **Application of Censored Regression Model to Factors Responsible for Extra-Marital Affairs.** *Rep Opinion* 2018;10(9):30-36]. ISSN 1553-9873 (print); ISSN 2375-7205 (online). http://www.sciencepub.net/report. 5. doi:10.7537/marsroj100918.05.

Keywords: Censored Regression, chi –square, Maximum Likelihood Estimator and R^2

1.0 Introduction

The general single-equation linear regression model, which is the universal set containing simple (two-variable) regression and multiple regression as complementary subsets, may be represented as

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + \varepsilon \quad (1)$$

Where Y is the dependent variable; $X_1, X_2, X_3, \ldots, X_k$ are independent variables; β_0 and β_i are regression coefficients, representing the parameters of the model for a specific population; and ε is a stochastic disturbance- term which may be interpreted as resulting from the effect of unspecified independent variables and/or a totally random element in the relationship specified.

The OLS estimate of the parameters is

$(\hat{\beta}) = (X'X)^{-1}X'Y$ (2)

Assume y and $x_{1i}...x_{ki}$ can be observed for every case in this study is censored, then the unknown values of $\beta's\ and\ \sigma$ in a regression model can be estimated from the observed data using ordinary least square method (OLS) but the OLS regression will treat the whole values as an actual and not as the upper limit of the affairs (y). A limitation of this approach is that when the variable is censored, OLS provide inconsistent estimates of the parameters, meaning that the coefficients from the analysis will not necessarily approach the "true" population parameters as the sample size increases.

1.1 Censored Regression

The Tobit is another name for censored regression which is a statistical model that describes the relationship between non-negative dependent variables y_i and an independent variable (or vector) x_i. The Tobit model can be described in terms of a latent variable y^*. Suppose, however that y_i^* is observed if $y_i^* > 0$ and is not observed. If $y_i^* \leq 0$. The observed y_i will be defined as

$$y_i = \begin{cases} y_i^* = \beta x_i + \varepsilon_i \ if\ y_i^* > 0 \\ 0 \ if\ y_i^* \leq 0 \end{cases} \quad \varepsilon_i \sim IIDN\ (0, \sigma^2) \quad (3)$$

The censored regression model involves some observation on y_i^* (those for which $y_i^* \leq 0$) are censored. The objective is to estimate the parameters β and σ. It is also known as a censored regression model which is design to estimate linear relationships between variables when there is either left- or right-censoring in the dependent variable.

In economics, such a model was first suggested in a pioneering work by Tobin [1958]. He analysed household expenditure on durable goods using a regression model which took account of the fact that the expenditure (the dependent variable of his regression model) cannot be negative. Tobin called the model, the model of limited dependent variables.

1.2 Overview of Extra-marital Affair

An affair is a sexual relationship, romantic friendship, or passion attachment between two people without the other spouse knowing. Affair may also describe part of an agreement within an open marriage or open relationship, such as swinging, dating, or polyamory, in which some forms of sex with one's non-primary partner (s) are permitted and other forms are not. Extramarital affair is described as the emotional adultery that culminates in physical sex or a long-term love affair. Kenkel (1985) highlighted different kinds of extramarital affairs which are: One night affair, Emotional sexual affair and Sexual addict affair. Extramarital affair could be caused as a result of something the couple did or something they ought to do but fail to do them. There are several factors responsible for extramarital affair which include; material and psychological deprivation, break in communication, sexual incompatibility, unemployment, childlessness, differences in interest, age, at marriage etc. Akinbodunse (1996) work on determinant of marital conflict of couples in Ondo; the findings revealed that unemployment and financial problems, of the husband has a significant role on the determination of marital conflict which is just an aspect on the causes of marital conflict. Honoree, khan and powell (2002) identify extramarital affair in five phases which are: Inception phase, Recovery phase, Pre-discovery phase, Discovery phase, Resolution phase.

The Tobit model is a statistical model proposed by James Tobin (1958) to describe the relationship between non-negative dependent variables and an independent variable (or vector). This paper described some of the assumptions of censored regression model and the interpretation of model and also suggested that maximum likelihood estimator is the best for the model when data is normally distributed. Takeshi Amemiya (1979) has proven that the maximum likelihood estimator suggested by Tobin for this model is consistent. Variation of the model can be produced by changing where and when censoring occurs. He classifies this variation in to five categories (Tobin type one -Tobin type five) and provide a general formula to obtain consistent likelihood estimators for these and other variations of the Tobit model. He provide an explicit, unified frame work for maximum likelihood estimation with –dimensional censored variables. Starting from the specific of the censoring problem and the distribution of the latent variables, it explain how to find an objective function such that its maximiser has all the asymptotic properties of maximum likelihood estimator. Karl. G Joreskog (2002) said a censored variable has a large fraction of observations at the minimum or maximum because the censored variable is not over its entire range. Ordinary estimates of mean and variance of a censored variable will be biased. He also explain how maximum likelihood estimates can be obtain using PRELIS. It also examined the use of tobit model for analysing measures of health status, the work examine relationships between determinants of health and measures of health status using censored regression.

Fair, R.C. (1978). Presented a journal on Theory of Extramarital Affair. This paper considers the determinant of leisure time spend in one particular type of activity with not household members: extramarital affairs. James L. Powell (October 1983) published Least Absolute Deviations Estimation for the Censo red regression model. In his paper he proposes an alternative to maximum likelihood estimation of the parameter of the censored regression (or censored 'Tobit') model if the data is non-normal. Hausman and Wise (1976 and1977) presented a journal on the estimation of earning equations on artificially-created, income truncated samples using Tobit type maximum likelihood estimation.

The aim of this research is the application censored regression model (Tobit) to factors that determine extramarital affair and make prediction based on the model fit.

2.0 Methodology

The censored regression model is a generalization of the standard Tobit model. The dependent variable can be either left-censored, right-censored, or both left-censored and right-censored, where the lower and / or limit of the dependent variable can be any number:

$$Y_i^* = X_i \beta + \varepsilon_1 \quad (4)$$

$$y_i^* = \begin{cases} a \; if \; y_i^* \leq a \\ y_i^* \; if \; a < y_i^* < b \\ b \; if \; y_i^* \geq b \end{cases} \quad (5)$$

Where

a = lower limit of the dependent variable.

b = upper limit of the dependent variable.

A censored variable has a large fraction of observation at the minimum or maximum. The censored variable is not observed over its entire range, ordinary estimates of the mean and variance of a censored variable will be biased. Green (2000) also observed Ordinary least square (OLS) estimates of its regression on a set of explanatory variables will also be biased. These estimates are not consistent, i.e., the bias does not become smaller when the sample size increases. Woolridge (2002) states that the maximum likelihood estimates are consistent, i.e. the bias is small in large samples.

The structural equation in this model is:

$y_i^* = \beta_o + \beta_1.gender + \beta_2.age + \beta_3.religious + \beta_4.yearsmarried + \beta_5.educatin + \beta_6.rating + error\ term$ (6)

$$y_i^* = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \varepsilon_i \quad (7)$$

Where $\varepsilon_i \sim N(0, \sigma^2)$. y_i^* is a latent variable that is censored. The observed y_i^* is defined by the following measurement equation

$$y_i^* = \begin{cases} y_i^* \ if\ y_i^* > 0 \\ 0 \ if\ y_i^* \le 0 \end{cases} \quad (8)$$

Estimation of Parameters Using Maximum Likelihood Estimator (MLE)

$y_i^* = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \varepsilon_i$ (9)

$\varepsilon_i \sim N(0, \sigma^2)$

$Y_i \sim N(\beta_o + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6, \sigma^2)$ (10)

The density function for the observations is

$$F = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{(-\frac{1}{2}(Y_i - \beta_o - \beta_1 x_1 - \beta_2 x_2 - \beta_3 x_3 - \beta_4 x_4 - \beta_5 x_5 - \beta_6 x_6)^2)}{\sigma^2}} \quad (11)$$

$$L((\beta_o, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6), \sigma^2) = \prod_{i=1}^{n} \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} e^{(-\frac{1}{2\sigma^2}(Y_i - \beta_o - \beta_1 x_1 - \beta_2 x_2 - \beta_3 x_3 - \beta_4 x_4 - \beta_5 x_5 - \beta_6 x_6)^2}$$

$$= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}((Y_i - \beta_o - \beta_1 x_1 - \beta_2 x_2 - \beta_3 x_3 - \beta_4 x_4 - \beta_5 x_5 - \beta_6 x_6)^2} \quad (12)$$

The log function for the model is:

$$log_e^L = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\text{Log}(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_o - \beta_1 x_1 - \beta_2 x_2 - \beta_3 x_3 - \beta_4 x_4 - \beta_5 x_5 - \beta_6 x_6)^2 \quad (13)$$

Differentiate with respect to $(\beta_o, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6), \sigma^2$

$$\frac{\partial log_e^L}{\partial(\beta_o)} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_o - \beta_1 x_1 - \beta_2 x_2 - \beta_3 x_3 - \beta_4 x_4 - \beta_5 x_5 - \beta_6 x_6)(-1) \quad (14)$$

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_o - \beta_1 x_1 - \beta_2 x_2 - \beta_3 x_3 - \beta_4 x_4 - \beta_5 x_5 - \beta_6 x_6) = 0 \quad (15)$$

$$\sum_{i=1}^{n}(Y_i - \beta_o - \beta_1 x_1 - \beta_2 x_2 - \beta_3 x_3 - \beta_4 x_4 - \beta_5 x_5 - \beta_6 x_6) = 0 \quad (16)$$

$$\sum_{i=1}^{n} y_i - n\beta_o - \beta_1 \sum_{i=1}^{n} x_1 - \beta_2 x_2 - \beta_3 \sum_{i=1}^{n} x_3 - \beta_4 \sum_{i=1}^{n} x_4 - \beta_5 \sum_{i=1}^{n} x_5 - \beta_6 \sum_{i=1}^{n} x_6 = 0 \quad (17)$$

$$\sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_1 - \beta_2 x_2 - \beta_3 \sum_{i=1}^{n} x_3 - \beta_4 \sum_{i=1}^{n} x_4 - \beta_5 \sum_{i=1}^{n} x_5 - \beta_6 \sum_{i=1}^{n} x_6 = n\beta_o \quad (18)$$

$$\beta_o = \frac{\sum_{i=1}^{n} y_i}{n} - \frac{\beta_1 \sum_{i=1}^{n} x_1}{n} - \frac{\beta_2 \sum_{i=1}^{n} x_2}{n} - \frac{\beta_3 \sum_{i=1}^{n} x_3}{n} - \frac{\beta_4 \sum_{i=1}^{n} x_4}{n} - \frac{\beta_5 \sum_{i=1}^{n} x_5}{n} - \frac{\beta_6 \sum_{i=1}^{n} x_6}{n} \quad (19)$$

Others value of the parameters are obtained through the same process of MLE.

3.0 Results and Discussion

The sample data is provided in appendix A

Table 1: Criteria for Evaluating Factors Affecting Extra-Marital Affairs

	Affairs	Age	Gender	Years married	Religion	Education	Rating
Valid	260	260	260	260	260	260	260
Missing	0	0	0	0	0	0	0
Mean	2.0308	43.1154	1.2808	11.9365	1.4692	14.8077	1.3577
Std. Error of Mean	.10650	.39668	.02792	.38980	.03101	.04598	.04287
Std. Deviation	1.71722	6.39635	.45024	6.28527	.50001	.74145	.69119
Variance	2.949	40.913	.203	39.505	.250	.550	.478

From the table 1 above, it can be inferred that gender has the lowest variance which means that gender has the largest influence on extramarital affairs. Also with the value of variance, it means that gender, religion, age in marriage has influence on extramarital affairs.

Table 2: Distribution of Number of Affairs According to Genders

number of affairs	gender		Total
	male	female	
NONE	82	60	142
ONE	50	13	63
TWICE	23	0	23
3-6TIMES	13	0	13
7-10TIMES	1	0	1
WEEKLY	10	0	10
MONTHLY	8	0	8
Total	187	73	260

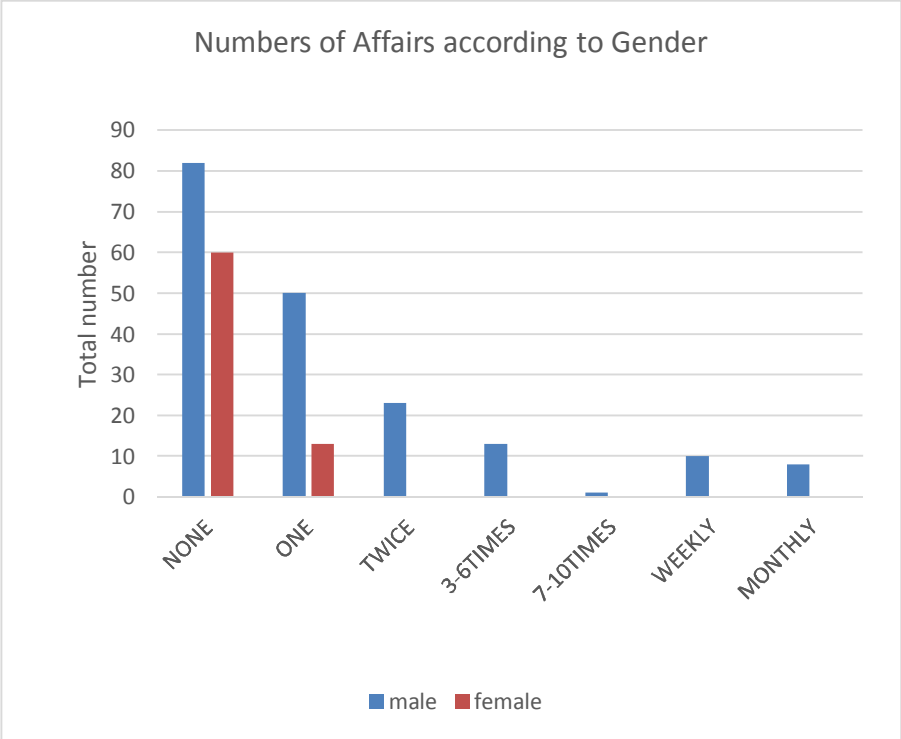

Figure 1: Number of Affairs according to Gender

Table 3: Distribution of Number of Affairs According to Religion

Number of affairs	Religion		Total
	Christianity	Islam	
None	102	40	142
One	49	14	63
Twice	17	6	23
3-6times	8	5	13
7-10times	0	1	1
Weekly	8	2	10
Monthly	7	1	8
Total	191	69	260

From the table 3 above it can be inferred that Christianity has the highest frequency of marital affairs than Islam.

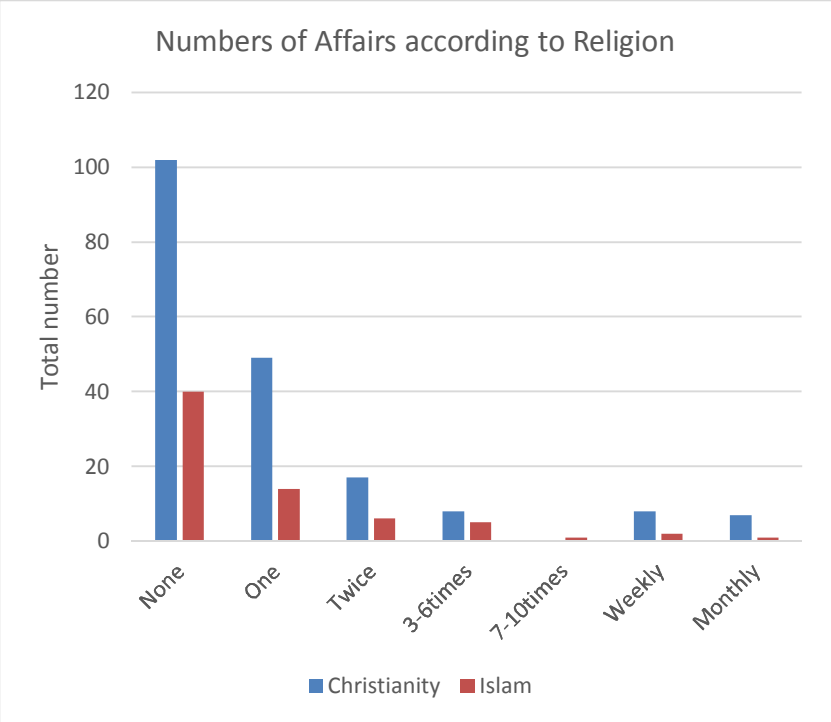

Figure 2: Numbers of Affairs according to Religion

3.1: Analysis on Factors that Contribute to Extra Marital Affairs Using Censored Regression Model.

H_0: there is no significant different among factors that contribute to extramarital affairs.

H_1: there is significant different among factors that contribute to extramarital affairs.

Decision rule: reject H_0 if p-value is less than 0.05 otherwise do not reject.

Table 4

Variable	Coefficient	Standard Error	Z-Statistic	P-value
C	10.69902	5.066339	2.111785	0.0347
EDUCATION	-0.521391	0.320057	-1.629058	0.1033
RELIGION	0.540626	0.056349	-2.108949	0.0049
RATING	0.927884	0.153833	6.031778	0.0002
GENDER	2.188614	0.040012	-4.052896	0.0001
AGE	-0.028957	0.276092	0.380552	0.7035
YEARSMARRIED	-0.038507	0.74690	-0.515555	0.6062

R-SQUARED = 0.656423

It can also be deduced with the p value < 0.05, that religion, rating and gender are the factors responsible for extramarital affairs. On the other hand, the p value is greater than 0.05 that means the numbers of year married, education qualifications and age of the respondent does not contribute to extramarital affairs. From the value of the R-squared above it can be deduce that the fit is good for the model and conclude that there is significant different between the factors that affect extramarital affairs

The fitted model is

Marital Affairs = 10.69902 – 0.521391 (Education) + 0.540626 (Religion) + 0.927884 (Rating) +2.188614 (Gender) – 0.028957(Age) −0.038507(Year Married)

Conclusion

Using the available data on questionnaire design, a predictive model was developed to estimate the factor responsible for number of extramarital affair. Censored regression model, relating education, age, rating, religion, gender and number of years married information on married staff in the school of sciences and school of agriculture was estimated. We conclude that rating, gender and religion has influence on the number of extramarital affairs with the p-value < 0.05.

References

1. Takeshi, Amemiya, (1979). Regression analysis when the dependent variable is truncated normal. Econometrica, 41,997-1016.
2. B. Honoree, S. Khan & J. L. Powell (2002). Quartile regression under random censoring. Journal of Econometrics, 109 67-105.
3. Benjamin Chijioke Asogwa, Victoria Ada Okwoche, Joseph Chinedu Umeh (2012) Analyzing the Determinants of Poverty Severity among Rural Farmers in Nigeria: A Censored Regression Model Approach. American international journal of contemporary research.
4. E. King, J. D. Hart & T. E. Wehrly (1991). Testing the equality of two regression curves using linear smoothers. Statistics & Probability Letters 12, 239-247.
5. Fair, R. (1978). A theory of extramarital affairs. Journal of political Economy, 86, 45-61.
6. Greene, W.H. (2000). Econometric Analysis. Fourth Edition. London: Prentice Hall International.
7. J. L. Powell (1984). Least absolute deviations estimation for the censored regression model. Journal of Econometrics, 25, 303-325.
8. Karl, G. Joreskog (2002). structural equation modelling with ordinal variables using LISREL. Available at http://www.ssicentral.com/lisrel/corner.htm
9. L. C. Zhao (2004). linear hypothesis testing in censored regression models, University of Science and Technology of China.
10. McDonald, J. F. and Moffitt, R. A. (1980). The Uses of Tobit Analysis. The Review of Economics and Statistics. Vol 62(2): 318-321.
11. Tobin, J. (1958). Estimation of relationships for limited dependent variables. Econometrica, 26, 24-36.
12. Wooldridge, Jeffrey (2002). Econometric Analysis of Cross Section and Panel Data. Cambridge: MIT Press.

Appendix A: Sample Data

S/N	Affairs	gender	age	Years married	religion	Education
1	1	1	33	2.5	1	13
2	1	2	33	2.5	1	14
3	1	1	33	2.5	1	13
4	1	1	38	7.5	1	15
5	1	1	38	2.5	1	15
6	1	1	38	7.5	1	15
7	1	1	38	7.5	1	14
8	1	1	38	2.5	1	14
9	3	1	38	7.5	1	14
10	3	1	38	7.5	1	14
11	3	1	38	7.5	1	14
12	1	1	38	7.5	1	14
13	7	1	38	2.5	1	15
14	1	1	38	7.5	1	14
15	1	1	48	13	1	14
16	1	1	38	7.5	1	14
17	1	2	43	7.5	1	14
18	1	2	48	13	2	14
19	1	2	48	18	2	15
20	1	2	48	18	2	14
21	1	2	48	18	1	14
22	1	2	48	13	1	14
23	1	2	48	13	1	14
24	2	2	48	13	1	13
25	2	2	48	18	1	13
26	2	2	48	13	1	14
27	2	2	48	13	1	15
28	2	2	48	13	1	15
29	2	1	48	13	1	15

S/N	Affairs	gender	age	Years married	religion	Education
30	2	2	48	13	1	15
31	2	1	48	13	1	16
32	2	2	48	13	1	15
33	2	1	53	18	1	14
34	2	1	43	13	1	16
35	1	1	53	23	1	16
36	1	1	53	23	1	15
37	1	2	33	2.5	1	14
38	1	2	33	2.5	1	14
39	1	2	33	2.5	1	14
40	2	1	33	2.5	1	15
41	3	1	33	2.5	1	14
42	4	1	33	2.5	1	14
43	4	1	33	2.5	1	14
44	2	1	33	2.5	1	14
45	4	1	33	2.5	1	14
46	1	1	33	2.5	1	14
47	8	1	33	2.5	1	14
48	1	2	33	2.5	1	14
49	2	2	33	2.5	1	15
50	1	2	38	7.5	1	14
51	1	2	43	13	1	14
52	1	2	33	2.5	1	15
53	2	1	48	13	1	16
54	8	1	43	13	1	15
55	2	1	33	2.5	1	14
56	2	1	43	7.5	2	15
57	2	1	48	13	2	14
58	8	1	53	18	2	15
59	4	1	53	23	2	16
60	8	1	53	23	2	15
61	1	1	53	23	2	16
62	3	1	53	23	2	16
63	7	1	53	23	2	16
64	8	1	53	23	2	16
65	1	1	53	23	2	15
66	3	1	53	23	2	16
67	8	1	53	23	2	15
68	8	1	48	13	2	16
69	2	1	48	13	2	16
70	1	2	48	13	2	15

9/25/2018