Using a Poisson Regression Model to analyze woman’s Labor Force data

Reem Ismail Elsybaey* (reemismail_2008@yahoo.com)

CAPMAS (Central Agency for Public Mobilization and Statistics), CAIRO, EGYPT

Supervised by:

Prof Dr: Amany Moussa**
Dr: Yasmin Ibrahim***

Keywords: Working Woman, Labor Force Survey, Deviance.

1. INTRODUCTION

Women’s participation in the labor force faced by a lot of obstacles; perhaps the most important are conditions and social beliefs, gender discrimination between men and women, which represents one of the structural problems in the labor market in Egypt. Discrimination appears in partitioning of jobs and in wages levels which reduce the motivation for woman’s participation in the labor force.

2. IMPORTANCE OF STUDY

According to many studies there exists a relationship between the number of participated women in the labor force and some other determinants like the number of educated, married women and place of residence. This paper will apply a Poisson Regression Model for the number of women participation in labor force in Upper Egypt. It will identify the determinants that have an impact on women's participation in the labor market.

3. METHODS

This study will use data of (labor force survey for the years 2011 and 2013) due to available data on workers and the unemployed, whether male or female. It will apply a poisson regression model to test the significance of some factors and it will use two measures of goodness-of-fit like deviance and AIC. In statistics, poisson regression is a form of regression analysis used to model count data and contingency tables. It assumes the response variable Y has a Poisson distribution, and assumes the logarithm of its expected value can be modelled by a linear combination of unknown parameters.
4. RESULTS

Figure 1. Ratio of female to male labor force participation (in %)

Figure (1) shows that the Ratio of female to male labor force participation was 34.5 in 1990, in 2002 it decreased to be 25.5, in 2008 it reached 30.8 and in 2012 it was 31.6.

Figure (2). The Residuals against the predicted value of mean of response

Figure (2) shows that in both years 2011 and 2013 there is over-dispersion in the data and the variance is too large. So the assumption in Poisson regression that at each level of the covariates the number of cases has variance equal to the mean is not violated or exist.
Table (1): The descriptive statistics for the data in 2011 and 2013

<table>
<thead>
<tr>
<th>Variables</th>
<th>2011</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of observations</td>
<td>Mean</td>
</tr>
<tr>
<td>Number of working women (Work)</td>
<td>44</td>
<td>136.30</td>
</tr>
<tr>
<td>Number of educated women</td>
<td>44</td>
<td>311.84</td>
</tr>
<tr>
<td>Number of married women</td>
<td>44</td>
<td>444.09</td>
</tr>
<tr>
<td>Number of women in age group (15-64)</td>
<td>44</td>
<td>693.98</td>
</tr>
</tbody>
</table>

In table (1) each variable has 44 valid observations in 2011 and 42 observations in 2013 also their distributions seem quite reasonable. The results show that the variance is greater than the mean which means that there is over dispersion in the data. Consequently, the estimates in Poisson Regression model are still consistent but are inefficient, which leads to the invalidation of inference based on the estimated standard errors. In this case the Negative binomial regression can be used for over-dispersed count data that is when the conditional variance exceeds the conditional mean \([4]\).

Table (2): Poisson Regression results for the data of 2011 and 2013:

<table>
<thead>
<tr>
<th>Exp. Variables</th>
<th>2011</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
</tr>
<tr>
<td>Intercept</td>
<td>3.705</td>
<td>0.032</td>
</tr>
<tr>
<td>Ur_Ru = 0 or 1</td>
<td>0.286</td>
<td>0.035</td>
</tr>
<tr>
<td>Number of educated women (Edu)</td>
<td>-0.002</td>
<td>0.000</td>
</tr>
<tr>
<td>Number of women in age group (15-64)</td>
<td>-0.002</td>
<td>0.000</td>
</tr>
<tr>
<td>Number of married (women (Mar)</td>
<td>0.006</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Recall the form of the model equation in years 2011 and 2013:

**In 2011:**
\[
\log(\text{Work}) = 3.705 + 0.286 (\text{Ur}_R\text{u}) - 0.002 (\text{Edu}) - 0.002 (\text{Age}) + 0.006 (\text{Mar}).
\]
This implies that:
\[
\text{Work} = \exp [3.705 + 0.286 (\text{Ur}_R\text{u}) - 0.002 (\text{Edu}) - 0.002 (\text{Age}) + 0.006 (\text{Mar})].
\]

**In 2013:**
\[
\log(\text{Work}) = 3.623 + 0.153 (\text{Ur}_R\text{u}) - 0.003 (\text{Edu}) + 0.004 (\text{Age}) - 0.002 (\text{Mar}).
\]
This implies that:
\[
\text{Work} = \exp [3.623 + 0.153 (\text{Ur}_R\text{u}) - 0.003 (\text{Edu}) + 0.004 (\text{Age}) - 0.002 (\text{Mar})].
\]

To assess the fit of the model, the goodness-of-fit chi-squared test is used where the deviance statistic can be approximated by a chi-square distribution when \( \mu \)'s are large. The deviance in 2011 was \( 1009.530 \) as Chi-square distributed with the model degrees of freedom \( 39 \), while in 2013 the deviance was equal to \( 482.387 \) with the model degrees of freedom \( 37 \).

The results of the data of 2011 and 2013 showed that all variables are significant and have effects on the number of participated women in labor force in Upper Egypt.

According to the standard errors in the data of the two years 2011 and 2013 it was too small and this shows that there is a high precision.

5. **CONCLUSIONS**

- The test of model effects showed that the number of educated and married females have a significant effect on the number of working women in Upper Egypt in both years 2011 and 2013.
- According to over dispersion the data is over dispersed, as the variance of the response variable is too large and not equal to the mean in both years 2011 and 2013.
- We tested over dispersion in the Poisson regression model using the ratio of the sum of Pearson chi square over the number of degrees of freedom (chi (2)/df). The results show that in 2011 the (chi (2)/df) = 23.109 and in 2013 (chi (2)/df) =12.009 which confirmed over dispersion.
- Negative binomial regression is for modeling count variables, usually for over-dispersed count outcome variables. Bouche, (2009) discussed in a paper different methods that exist to solve the problem of underestimating variance in the Poisson regression model when over dispersion is present [5].

6. **REFERENCES**


