Logistic Regression Analysis of Disabled Employee Data

Berna YAZICI[a], Betül KAN[a], Yener ŞİŞMAN[b], Fatma KOCABAŞ[b]

[a] Faculty of Science, Department of Statistics, Anadolu University, Eskişehir, Turkey
[b]Faculty of Economics and Administrative Sciences, Department of Labor Economics and Industrial Relations, Anadolu University, Eskişehir, Turkey

A survey is conducted in order to determine the demographic structure and job satisfaction of disabled people in working environment. The population is defined as disabled people who are working for the companies which are obliged to employ disabled people in Eskişehir, Turkey. It is thought that almost 1,000 disabled employees are working in private and public sector. 421 of those are reached and asked to complete a questionnaire. In this paper the demographic structure of disabled employees is given. Besides, logistic regression models are constructed; binary logistic regression model in order to find out whether or not they are meeting any handicaps at their working lives, ordinal logistic regression model for job satisfaction of disabled employees with their works and nominal logistic regression model for the task types of disabled people at work. The models are given and interpreted.

Keywords: Disability; Employment of Disabled People; Binary Logistic Regression; Nominal Logistic Regression; Ordinal Logistic Regression.

Introduction

The employment and problems of disabled people in working life have always been of interest. Doyle, (1995) examined the role that the law might play insuring the employment rights of disabled people in Britain by comparing the legislative frameworks of Australia, Britain, Canada, the European Union and the United States. Barnes, (2002) raised issues of disabled people's employment and living standards from a UK perspective, and addressed them through and analysis of policies in three West European countries; Britain, Germany and Sweden. In 2005 UK Parliament passed of the Disability Discrimination Act. Roulstone et al., (2006) focused on the employment provisions of the 2005 Act and challenges of applying a barriers approach to a disability employment monitoring schemes. Angus et al., (2006) considered attitudes of human resource managers towards the employment and employability of disabled people in UK accounting firms. They identified that specific impairments which firms see as impediments to employment and critically assesses firms' perceptions of the employability of disabled people. Roggero et al., (2006) studied the results of an e-discussion hosted by the World Bank with 3900 contacts in order to see the employment disabled youth in developed and developing countries. Bishop et al., (2008) focused on the notions of disability and explained how UK legislative regimes were designed to support and sustain such classificatory regimes. Shier et al., (2009) examined the barriers to employment as experienced by disabled people via interviewing 56 disabled individuals.

In Turkey there have been two survey studies conducted. Aydın, (1991) examined the problems of disabled employers who have been working in Zonguldak (a city famous for coal mines in north part of Turkey). Yılmaz, (2004) examined the problems of disabled people in working life and the factors affect them.

This study is regarding a survey that is conducted in order to model the disabled peoples' satisfaction in working life and model their task types. Firstly, binary, nominal and ordinal logistic regression models are reviewed, then the findings of the survey, demographic structure of the respondents, the models for this data set and the interpretations of the models are given.

The Logistic Regression Model

Logistic regression is used when the response variable is binary, nominal and ordinal. For the binary response variable case, the model would take the following form;

$$y_i = \mathbf{x_i}'\boldsymbol{\beta} + \varepsilon_i \qquad (1)$$

where $\mathbf{x_i}' = [1, x_{i1}, x_{i2}, ..., x_{ik}]$, $\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2, ..., \beta_k]$ and the response y_i can only take on the values of 0 or 1.

y_i=1, $P(y_i = 1) = \pi_i$ (usually takes the value of 1 when $P(y_i \geq 0{,}5)$ in application)

y_i=0, $P(y_i = 0) = \pi_i$ (usually takes the value of 0 when $P(y_i < 0{,}5)$ in application)

Since the response is binary, an s-shaped curve is used and this function is called logistic or logit. The form of logit function is given in Equation (2):

$$E(y) = \pi = \frac{e^{g(\mathbf{x})}}{1+e^{g(\mathbf{x})}} = \frac{1}{1+e^{-g(\mathbf{x})}} = \hat{y} \qquad (2)$$

Where g(**x**)= $\mathbf{x}'\boldsymbol{\beta}$. The g(**x**) is defined as the linear predictor. g(**x**) can also be written as in Equation (3):

$$g(\mathbf{x}) = \ln\frac{\pi}{1-\pi} \qquad (3)$$

The binary case can be generalized for multi-level response case. When there are more than two nominal response categories, logistic regression fits a model used generalized logits. The generalized logit is defined as follows:

$$h(x)_j = \log\left[\frac{\pi_j}{\pi_r}\right] \qquad (4)$$

for j=1, 2, … , (r-1) nominal response categories. A logit is formed for the probability of each succeeding category over the last response category (Lawson and Montgomery, 2006). In other words, generalized logits are computed that are based on each response level compared with one designated reference response levels. The generalized logits for a five level response would be

$$h(x)_1 = \log\left[\frac{\pi_1}{\pi_5}\right],\ h(x)_2 = \log\left[\frac{\pi_2}{\pi_5}\right],\ h(x)_3 = \log\left[\frac{\pi_3}{\pi_5}\right],\ h(x)_4 = \log\left[\frac{\pi_4}{\pi_5}\right]$$

Owing to the way in which the logits are calculated, the reference category becomes the category against which all the other responses are compared.

The generalized linear predictor model is given using Equation (5):

$$\hat{g}(x)_k = \beta_{0k} + \mathbf{x_i}'\boldsymbol{\beta}_k \qquad (5)$$

where k is the index of the logits.

The third type is the ordinal logistic regression which is modeled for the orderly measured response, from the lowest to the highest or from the highest to the lowest. In ordinal regression cumulative logits are computed that are based on cumulative probabilities of the response levels. This approach known as the proportional odds model, takes the rank ordering of the response into account. With this model the probability of an equal or smaller response $Y \le k$, is compared with the probability of a larger response, Y>k;

$$h_k(x) = \ln\left[\frac{P(Y \le k|\mathbf{x})}{P(Y > k|\mathbf{x})}\right] \qquad (6)$$

where k is the rank of the ordinal categories. The predicted values are computed in the same manner in nominal logistic regression;

$$\hat{g}(x)_k = \beta_{0k} + \mathbf{x}'\boldsymbol{\beta}_k \qquad (7)$$

The models are fitted with the same set of slope parameters but different intercepts for each logit (Lawson and Montgomery, 2006).

Application

Data was obtained with the survey which was conducted over 6 months. The survey was conducted in Eskişehir, a large industrial Turkish city. Demographic structure of the data is summarized and given in Table 1.

Table 1. Demographic structure of disabled employees

Variable of Interest		Response frequencies	Percent
Gender			
	Male	341	81.0%
	Female	75	17.8%
Marital Status			
	Single	160	38.0%
	Married	232	55.1%
	Divorced	21	5.0%
	Widowed	5	1.2%
	Separated	3	0.7%
Education			
	Secondary school degree	185	44.5%
	High school degree	181	43.6%
	Bachelor degree	49	11.8%
Type of Disability			
	Orthopedically	164	39.0%
	Hearing	84	20.0%
	Sight	71	16.9%
	Mental	25	5.9%
	Speech	6	1.2%
	Other	57	13.5%

Appearance Time of Disability			
	Congenital	196	47.6%
	Subsequent	216	52.4%

The ages of the employees range between 19 and 57 with mean 31.74. Only 4.3% of the respondents live alone. The others live within the family of 1 to 10 people with mean 3.5.

Binary Logistic Regression Results

Disabled employees are asked if they meet handicaps at work or not, and the behaviors of their colleagues and employers.

Table 2. Response variable's distribution in Binary logistic regression

Meeting Handicaps	Frequency	Marginal Percentage
Yes, I meet with handicaps	94	23.2%
No, I don't meet with handicaps	311	76.8%

16 of the disabled employees did not reply that question.

Table 3. Predictor variable's distribution in Binary logistic regression

Task Type	Frequency	Marginal Percentage
Domineering	24	5.9%
Extremely protecting	20	4.9%
With love and respectful	305	74.8%
Other	59	14.5%

There are 13 respondents who did not answer the question about the behavior.

Table 4 gives the goodness of fit for binary logistic regression. G denotes the significance of the model and $p<0.05$ shows that the null hypothesis that the estimates of the β coefficients are equal to zero should be rejected, the model is significant.

Table 4. Model summary statistics for Binary logistic regression

Method	Statistic	Degrees of freedom	P-value
G	27.265	3	0.000

The significance of individual parameter estimates are obtained using Wald test and the results are given in Table 5. It can be concluded that the coefficient for protecting behavior is not significant while the others are significant.

Table 5. Logistic regression table for binary data

Predictor	Coefficient	Standard error of coefficient	Z-Value	P-Value	Odds Ratio	95% CI Lower	95% CI Upper
Intercept	0.566	0.273	2.073	0.038			
Behavior (domin.)	-1.340	0.564	-2.375	0.018	0.262	0.087	0.791
Behavior (protect.)	1.108	0.686	1.615	0.106	3.027	0.789	11.611
Behavior (love-res.)	0.928	0.311	2.983	0.003	2.528	1.374	4.654

$$\hat{g}(x) = 0.566 - 1.340(\text{Behavior-domin.}) + 1.108(\text{Behavior-protect}) + 0.928(\text{Behavior-love})$$

Table 6 gives the predicted probabilities using Equation 2.

Table 6. Predicted probabilities for Binary logistic regression

Behavior	Probability of meeting with a handicap	Probability of not meeting with a handicap
Domineering	0.68	0.32
Protecting	0.16	0.84
Love-respectful	0.18	0.82
Other	0.36	0.64

According to binary logistic regression results, if disabled person's collogues behave him/her domineering, he/she will most probably (68%) meet a handicap at work.

Nominal Logistic Regression Results

Disabled employees are asked about their task type at their work, education, gender and appearance time of disability. The distributions of those variables are given in Table 7 and Table 8.

Table 7. Response variable's distribution in Nominal logistic regression

Task Type	Frequency	Marginal Percentage
Production sector employee	255	66.6%
Service sector employee	74	19.3%
Office employee	54	14.1%

There are 38 respondents who did not answer all those 4 questions. There are 38 missing values at this part of the study.

Table 8. Predictor variables' distribution in Nominal logistic regression

Predictor	Frequency	Marginal Percentage
Education		
Secondary school degree	173	45.2%
High school degree	167	43.6%
Bachelor degree	43	11.2%
Gender		
Male	315	82.2%
Female	68	17.8%
Appearance Time of Disability		
Congenital	177	46.2%
Subsequent	206	53.8%

Table 9 gives the goodness of fit for nominal logistic regression. G denotes the significance of the model and $p<0.05$ shows that the null hypothesis that the estimates of the β coefficients are equal to zero should be rejected, the model is significant. On the other hand D denotes the deviance. For $p>0.05$, it can be concluded that the deviance is not significant and the model fit is good.

Table 9. Model summary statistics for Nominal logistic regression

Method	Statistic	Degrees of freedom	P-value
G	44.037	8	0.000
D	15.240	14	0.362

The significance of individual parameter estimates are obtained using Wald test and the results are given in Table 10. It can be concluded that while education and gender affecting the task type, the appearance time of disability does not affect the task type of the disabled.

Table 10. Logistic regression model for logit 1 (production sector employee/office employee)

Predictor	Coefficient	Standard error of coefficient	Z-Value	P-Value	Odds Ratio	95% CI Lower	95% CI Upper
Intercept	-0.369	0.416	-0.887	0.376			
Education (ss)	1.719	0.489	3.515	0.000	5.579	2.142	14.533
Education (hs)	0.519	0.421	1.233	0.217	1.681	0.737	3.837
Gender (Male)	1.250	0.365	3.425	0.001	3.491	1.708	7.135
ATD (Congenital)	0.389	0.328	1.186	0.236	1.476	0.775	2.811

Logit 1 compares working in production sector with office employee. The x value is 1 for secondary school graduates, 1 for males and 1 for who had the disability congenital. The odds ratio 5.579 indicates that the secondary school graduated disabled are 5.5 times more likely work in production sector versus work in office. Males are 3.54 times more likely work production sector versus females.

$\hat{g}(\mathbf{x})_1$ =-0.369+1.719(Education-ss)+0.519(Education-hs)+1.250(Gender-Male)+0.389(ATD-Congenital)

Table 11. Logistic regression model for logit 2 (service sector employee/office employee)

Predictor	Coefficient	Standard error of coefficient	Z-Value	P-Value	Odds Ratio	95% CI Lower	95% CI Upper
Intercept	-2.327	0.705	-3.301	0.001			
Education (ss)	2.806	0.744	3.772	0.000	16.541	3.852	71.038
Education (hs)	1.494	0.704	2.122	0.034	4.455	1.118	17.748
Gender (Male)	0.869	0.455	1.910	0.057	2.384	0.976	5.825
ATD (Congenital)	0.544	0.385	1.413	0.158	1.724	0.809	3.670

Logit 2 compares working in production sector with office employee. The x value in the model is 0 for secondary school graduates, 1 for high school graduates, 1 for males and 1 for who had disability congenital. The only difference with logit 1 is the intercept. For the secondary school graduates, the probability of working in service sector is 16.541 times that working as office employee.

$\hat{g}(\mathbf{x})_2$ = -2.327+2.806 (Education-ss) + 1.494 (Education-hs) + 0.869 (GenderMale) + 0.544 (ATD - Congenital)

Table 12 gives the predicted cell probabilities using Equation 2.

Table 12. Predicted cell probabilities for nominal logistic regression

ATD	Gender	Education	Probability of working in production sector	Probability of working in service sector
Congenital	Male	Secondary S.	0.72	0.24
		High S.	0.68	0.20
		Bachelor	0.72	0.08
	Female	Secondary S.	0.74	0.21
		High S.	0.67	0.17
		Bachelor	0.66	0.06
Sequential	Male	Secondary S.	0.60	0.29
		High S.	0.50	0.22
		Bachelor	0.47	0.08

Female	Secondary S.	0.60	0.25
	High S.	0.45	0.17
	Bachelor	0.38	0.05

According to the results, if disabled person is secondary school graduate, male and congenitally got his disability, he will most probably (72%) work in production sector.

Ordinal Logistic Regression Results

Disabled employees are asked whether they are satisfied with their job or not. The distribution about that response variable is given in Table 13.

Table 13. Response variable's distribution in Ordinal logistic regression

Satisfaction with the job	Frequency	Marginal Percentage
I am very satisfied and I never think changing my job	242	80.7%
I am not satisfied but I don't think changing my job	31	10.7%
I am not satisfied and I am planning changing my job	27	9.0%

Three predictor variables are used in order to model the job satisfaction; promotion, handicaps and education. The distributions of those variables are given in Table 14. There are 121 missing values for those 4 variables.

Table 14. Predictor variables' distribution in Ordinal logistic regression

Item	Frequency	Marginal Percentage
Promotion		
Employees who hope to promote	49	16.3%
Employees who don't think that they will promote	251	83.7%
Handicaps		
I meet with handicaps at work	69	23.0%
I don't with handicaps at work	231	77.0%
Education		
Secondary school degree	134	44.7%
High school degree	126	42.0%
Bachelor degree	40	13.3%

The model output is shown in Table 10. The proportional odds model produces one model that compares the satisfied ones with the ones who don't think changing their jobs, and one another model that compares the satisfied ones with the ones who think changing their jobs. Being hopeful about promoting in their jobs and meeting with handicaps in their jobs are the significant variables.

Table 15. Ordinal logistic regression table

Predictor	Coefficient	Standard Error	Z-value	P-value	Odds ratio	95%CI Lower	95%CI Upper
Intercept (1)	2.279	0.477	4.778	0.000		1.344	3.215
Intercept (2)	3.235	0.505	6.406	0.000		2.245	4.226
Promotion (hopeful)	0.736	0.359	2.050	0.040	2.088	0.033	1.439
Handicap (yes)	1.264	0.317	3.987	0.000	3.540	0.643	1.884
Education (ss)	0.194	0.513	0.378	0.706	1.214	-0.812	1.200
Education (hs)	0.466	0.504	0.925	0.355	1.594	-0.522	1.454

The probability of being satisfied with their jobs is two times less for the ones who are hopeful about promoting. The probability of being satisfied with their jobs is 3,5 times less for the ones who meet with handicaps in their jobs.

According to the first linear model given below, if the disabled is hopeful about promoting in their jobs, the first predictor will take 1, if they meet with handicap the second predictor will take 1, if they are graduated from high school the third predictor will take 0 and fourth will take 1. The same structure is valid for the second model, while the intercept changes.

$$\hat{g}(\mathbf{x})_1 = 2.279+0.736(\text{Promotion})+1.264(\text{Handicap})+0.194(\text{Educa.-ss})+0.466(\text{Educa.-hs})$$

$$\hat{g}(\mathbf{x})_2 = 3.235+0.736(\text{Promotion})+1.264(\text{Handicap})+0.194(\text{Educa.-ss})+0.466(\text{Educa.-hs})$$

Table 16. Model summary statistics for ordinal data for Ordinal logistic regression

Method	Statistic	Degrees of freedom	P-value
G	25.880	4	0.000
D	27.575	18	0.069

As given in Table 11, the model fit is adequate. Since the p value is 0.000, the null hypothesis that the slope parameters are all equal to zero is rejected. The deviance statistic (D) is a measure of the difference between the fitted model and the saturated model. Since the p value for D is 0.069>0.05 the model fit is adequate.

Table 17. Predicted cell probabilities for Ordinal logistic regression

Promotion	Handicap	Education	Probability of being very satisfied	Probability of not being satisfied
Hopeful	Meet hand.	Secondary S.	0.52	0.22
		High S.	0.45	0.23
		Bachelor	0.57	0.21
	Don’t meet hand.	Secondary S.	0.79	0.12
		High S.	0.75	0.14
		Bachelor	0.82	0.10
Not hopeful	Meet hand.	Secondary S.	0.69	0.16
		High S.	0.63	0.18
		Bachelor	0.73	0.14
	Don’t meet hand.	Secondary S.	0.89	0.06
		High S.	0.86	0.08
		Bachelor	0.91	0.05

The predicted probabilities are calculated using Equation 2 and the results are given in Table 17. The probability of being satisfied for the ones who are hopeful about promoting, who meet with handicaps in their jobs and who are graduated from high school is 0.52.

Conclusions

In this study the data set obtained by conducting a survey with 421 disabled employees. Firstly binary logistic regression model is constructed in order to find out the probabilities of meeting handicaps or not according to the behaviors of disabled employees’ colleagues or employers. Using the same data set nominal logistic regression model is constructed for the sectors which the disabled people might work regarding their gender, education and appearance time of their disability. Also their satisfaction from their jobs is modeled using ordinal logistic regression. For that part of the study, education, being hopeful or not about promoting in the job and meeting with a handicap or not in the job are taken as the predictor variables. It is seen that all models are significant.

Especially for the survey studies, the data sets are usually made up of the variables that are not normally distributed. Since those output and the predictor variables are not normally distributed, logistic regression is a useful modeling technique with its flexible and common use for such data sets.

Acknowledgments

The authors wish to thank Anadolu University Scientific Research Projects (Project No. 080601) for financial support.

References

Aydın, Y. 1991. Examining the employment problems of disabled peopled regarding the employment legislations and Zonguldak sample. Master thesis, Hacettepe University, Ankara.

Barnes, H. 2002. Working for a Living? Employment, Benefits and the Living Standards of Disabled People. *Journal of European Social Policy* 12, no.1: 86-87.

Bishop, M., and R. Boden. 2008. Disabling accounting. Critical Perspectives on Accounting 19: 1-16

Doyle, B. 1996. Disability, Discrimination and Equal Opportunities: a comparative study of the employment rights of disabled people. *Disability & Society* 11, no.2: 291-296.

Duff, A., J. Ferguson, and K. Gilmore. 2007. Issues concerning the employment and employability of disabled people in UK accounting firms: An analysis of the views of human resource managers as employment gatekeepers. *The British Accounting Review* 39, 15-38.

Lawson C., and D.C. Montgomery. 2006. Logistic Regression Anaysis of Customer Satisfaction Data. Quality and Reliability Engineering International 22: 971-984.

Roggero, P., R. Tarricone, M. Nicoli, and V. Mangiaterra. 2006. What do people think about disabled youth and employment in developed countries? Results from an e-discussion hosted by the World Bank. *Disability & Society* 21, no.6: 645-650.

Roulstone, A., and J. Warren. 2006. Applying a barriers approach to monitoring disabled people's employment: implications for the Disability Discrimination Act 2005. *Disability & Society* 21, no.2: 115-131.

Shier, M., J.R. Graham, J. Marion. 2009. Barriers to employment as experienced by disabled people: a qualitative analysis in Calgary and Regina, Canada. *Disability & Society* 24, no.1: 63-75.

Yılmaz, Z. 2004. The problems that the disabled face in working life and the factors that affect them. Master thesis, Hacettepe University, Ankara.