The Report Committee for Zhou Lu
Certifies that this is the approved version of the following report:

A Study of Courteous Behavior on the University of Texas Campus

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor: ______________________________
Chandler W. Stolp

____
Daniel A. Powers

A Study of Courteous Behavior on the University of Texas Campus

by

Zhou Lu, B.S., M.S., M.A.

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Statistics

The University of Texas at Austin

December 2010

Dedication

This Thesis is dedicated to my wife, son and parents who provided support with all their efforts.

Acknowledgements

Special thanks to the following persons who helped me in my graduate career through their own unique ways.

To my advisor, Dr. Chandler Stolp, for his kind helps on design, analysis and detailed STATA syntax.

To my committee, Dr. Daniel Powers, for his precious suggestions, time and patience.

And to my family for their support.

November 30, 2010

Abstract

A Study of Courteous Behavior on the University of Texas Campus

Zhou Lu, M.S.STAT.

The University of Texas at Austin, 2010

Supervisor: Chandler W. Stolp

This study focused on measuring courteous behavior on the University of Texas at Austin (UT) students on campus. This behavior was measured through analyzing various factors involved when a person opened the door for another. The goal was to determine which factors would significantly affect the probability that a person would hold a door for another. Three UT buildings with no automatic doors were selected (RLM, FAC and GRE), and 200 pairs of students at each location were observed to see whether they would open doors for others. These subjects were not disturbed during the data collection process. For each observation, the door holding conditions, genders, position (whether it was the one who opened the door or the recipient of this courteous gesture, abbreviated as recipient), distance between the person opening the door and the recipient, and the number of recipients were recorded.

Descriptive statistics and logistic regression were used to analyze the data. The results showed that the probability of people opening the doors for others was significantly affected by gender, position, distance between the person opening the door and the recipient, the number of recipients, and the interaction term between gender and position.

The study revealed that men had a slightly higher propensity of opening the doors for the recipients. The odds for men were a multiplicative factor of 1.09 of that for women on average, holding all other factors constant. However, women had much higher probability of having doors held open for them. The odds for men were a multiplicative factor of 0.55 of that for women on average, holding all other factors constant. In terms of the distance between the person opening the door and the recipient, for each meter increase in distance, the odds that the door would be held open would decrease by a multiplicative factor of 0.40 on average. Additionally, for each increase in number of recipients, the odds that the door would be held open would increase by a multiplicative factor of 1.32 on average.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Door holding etiquette as a mean of courteous behavior

According to social etiquette, opening the door for another person is commonly considered a sign of courteous behavior and good manners. This study was conducted to provide answers to the following questions: Which factors significantly affect the door holding conditions on the University of Texas at Austin (UT) campus? These factors include the gender of the person opening the door as well as person who accepted this gesture, the number of recipients of this courteous gesture (abbreviated as recipients), and the distance between these subjects.

The results of some studies suggested that for door opening etiquette, the essential factor may not be the gender, but rather, whoever is closer to the door. According to research by Wouters on relationship development between sex and manner for the period between 1890 and 2000 in the western world, chivalry and manners were “informalized” during the period as changes prompted in women’s identity and relationship between men and women [1]. Lakoff and Montgomery also brought out a theory stating that generally, women would be more polite than men [2, 3]. However, Gibson did not agree to the above theory and proved in her recent research that female cashiers neither treated their customer more politely than male, nor were they treated more politely by the customers [4].

In late 1990s, some journalists criticized the poor door holding manners of students from Massachusetts Institute of Technology (MIT). They claimed that MIT students failed to obey the general door rules on a daily basis almost everywhere on campus and at train stations. Today's door holding rules are more relaxed in comparison to those in the old days. However, whether the university students these days understand and obey the general door rules, either unconsciously or intentionally, may be an interesting debatable question.

This University of Texas at Austin is known for its large population of students. This population has a great diversity. The behaviors of UT students, to some extent, represent the manners of the entire society. Therefore, the study results based on the behaviors of UT students may then be extended to the whole nation through proper modifications.

1.2 SUMMARY OF CHAPTERS

Chapter 1, provides an introduction to the motivations for this study as well as the background information.

Chapter 2, focuses on the methodology used in the study, including introduction to logistic regression, how to collect the data, define the variables and perform the logistic regression.

Chapter 3, illustrates the results and conclusion, including analysis and output interpretation, advantages and disadvantages of the study, as well as potential improvements for future analysis.

1.3 REFERENCES

1. Wouters, C. 2004. *Sex and Manners: Female Emancipation in the West 1890-2000.* London: SAGE Publications.
2. Lakoff, R. T. 2004. *Language and Women's Place.* Rev. and expanded ed. New York; Oxford: Oxford University Press.
3. Montgomery, M. B. 1998. "Multiple Modals in LAGS and LAMSAS," In *From the Gulf States and Beyond: The Legacy of Lee Pederson and LAGS,* edited by Nunnally, T., 90-122, Tuscaloosa/London: University of Alabama Press.
4. Gibson, E. K. 2009. Gender, Polite Questions and the Fast-food Industry. *Griffith Working Papers in Pragmatics and Intercultural Communication* 2(1):1-17.

Chapter 2: Research Design and Methods Used

2.1 Logistic regression analysis

Logistic regression, also known as a logistic model or a logit model that is especially well suited to settings in which dependent variable is binary. It is used to predict the occurrence of an event by fitting an "S-shaped" logistic curve [1]. In order to gain answers to the questions asked at the beginning of this study, logistic regression is used to analyze the data.

2.1.1 Logistic function

Logistic function is the first step of understanding the logistic regression.

$$f(y|\boldsymbol{X}) = \frac{\boldsymbol{e}^{X\beta}}{1 + e^{X\beta}} = \frac{1}{1 + e^{-X\beta}} \qquad (1)$$

here f(y|**X**) can be explained as the probability of success, given a set of explanatory variables **X**, and it is constrained to lie between 0 and 1. The graph corresponding to the above function is shown as Figure 1. The advantage of this density function is that it is a continuous, monotonically increasing or decreasing function of the X.

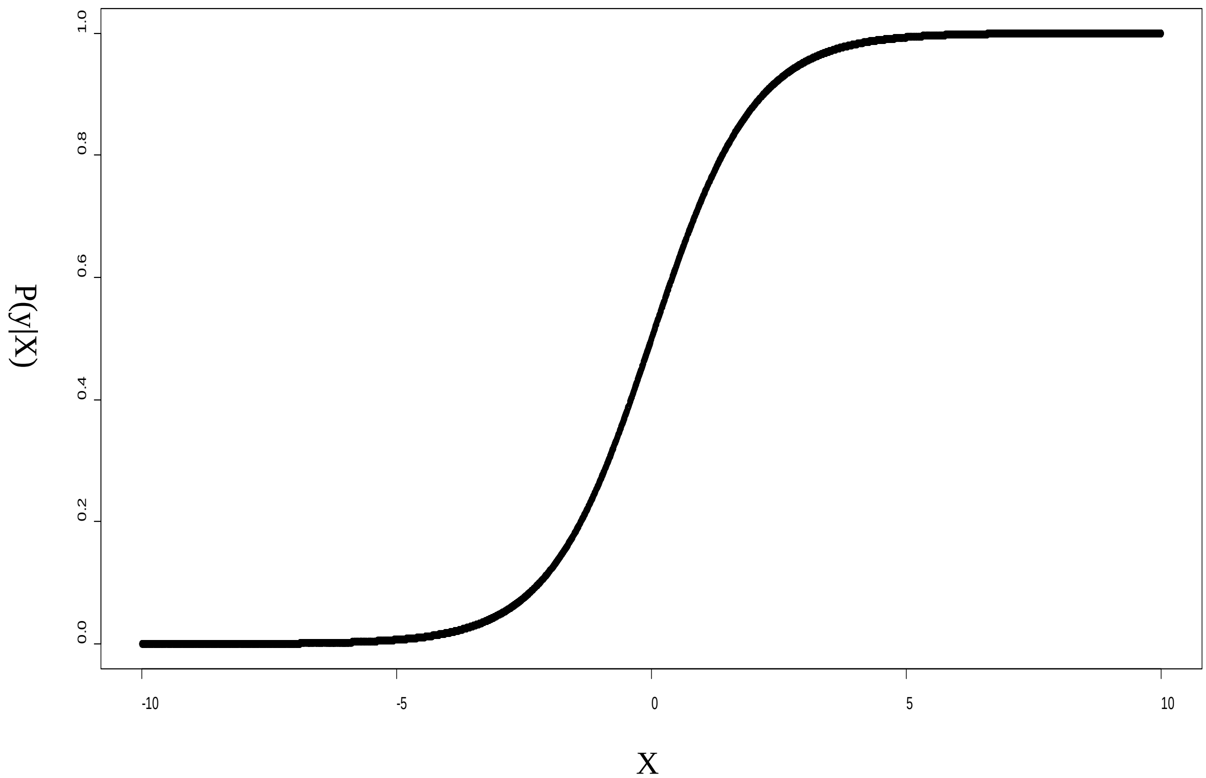

Figure 1: Logistic function. The vertical axis represents the probability of success and its value is constrained to be between 0 and 1.

2.1.2 Logistic regression

Logistic regression is well suited for the analysis of binary response variables. A binary variable is a variable with binary outcomes, such as success (1) or failure (0). The probability of success follows a Bernoulli distribution. It is known that the number of success (S) out of n trials with identical Bernoulli distribution (with probability of success fixed as p) is binomially distributed (noted as S~ B(n, p)).

The logistic density function reveals the probability of success, which is affected by a set of explanatory variables:

$$p(y|\boldsymbol{X}) = \frac{1}{1 + e^{-\boldsymbol{X}\beta}} \tag{2}$$

where $\boldsymbol{X}\beta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m$. The X's are explanatory variables, such as gender of the person, position of the person, distance between the people, as well as the number of recipients in this study.

By reforming equation (2), we obtain,

$$e^{X\beta} = \frac{p}{1 - p} \tag{3}$$

On the right side of equation (3) is the ratio of the probability of success and the probability of failure, which is defined as the odds of success. The natural log of the odds is known as a logit.

The logistic regression model is expressed more fully as:

$$\mathrm{logit(p)} = \ln\left(\frac{\mathrm{p}}{1 - \mathrm{p}}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m \tag{4}$$

In other words, probability of success can be predicted by

$$\hat{p} = \frac{1}{1 + e^{-(\widehat{\beta_0} + \widehat{\beta_1} X_1 + \widehat{\beta_2} X_2 + \cdots + \widehat{\beta_m} X_m)}} \tag{5}$$

where $\hat{p}$ is the predicted probability of success and $\hat{\beta}$s are sample estimates of the population coefficients βs [2-5].

The odds can be extended to odds-ratio, which describes the odds of success associated with one group against another [6]. For example, assuming two groups with probabilities of success p_1 and p_2, respectively, the corresponding odds are $\frac{p_1}{1-p_1}$ and $\frac{p_2}{1-p_2}$. Further assume $logit(p_1) = \beta_0 + \beta_1$ and $logit(p_2) = \beta_0 + \beta_2$, where β_0, β_1 and β_2 are parameters. The odds-ratio, upon measuring the odds of success in group 1 against group 2 is then,

$$odds\ ratio = \frac{\frac{p_1}{1-p_1}}{\frac{p_2}{1-p_2}} = \frac{e^{\beta_0+\beta_1}}{e^{\beta_0+\beta_2}} = e^{\beta_1-\beta_2}$$

2.1.3 Interpreting the coefficients

In the logistic model shown in equation (4), the parameters $\beta_0, \beta_1, \dots, \beta_m$ are estimated using the method of maximum likelihood [7]. Unlike the interpretation for the linear regression, $\widehat{\beta_1}$ ($\widehat{\beta_2} \dots, \widehat{\beta_m}$) is no longer interpreted as the expected increase in response variable y for a unit increase in the explanatory variable X_1 ($X_2, \dots, X_m$), holding all other variables in the model constant. Instead, it is interpreted as the expected increase in the logit, the natural log of odds, for a unit increase in explanatory variable X_1 ($X_2, \dots, X_m$), holding all other variables constant. The change in logit may not be easily explained since it involves natural log transformation. In practice, the coefficients have more meaningful interpretations when transformed exponentially. For example, instead of interpreting $\widehat{\beta_1}$ as the expected increase in logit $\ln(\frac{p}{1-p})$ for every unit increase in explanatory variable X_1, holding all other variables constant, $e^{\widehat{\beta_1}}$ is interpreted as the multiplicative change in the odds of success per unit increase in the explanatory variable X_1, holding all other variables constant. In other words, the odds will increase by $100 \times (e^{\widehat{\beta_1}} - 1)\%$, decrease if the effect is negative, for a unit increase in the explanatory variableX_1, holding all other variables constant.

2.1.4 Interactions between continuous or discrete explanatory variables

Interaction effects refer to how the impact of one explanatory variable on the response variable depends on the magnitude of another explanatory variable. Assuming the response variable y depends on explanatory variablesX_1, X_2, their interaction term X_1X_2, and a set of other variables denoted as X. In linear models, the expected value of y conditioned on X_1, X_2 and X is expressed as,

$$E[y|X_1, X_2, X] = \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + X\beta \qquad (7)$$

The interaction effect is indicated by β_{12}, the coefficient associate with the interaction term X_1X_2.

Norton and his colleagues point out that the interpretation of interaction effects is not as straight forward in nonlinear models such as logit and probit as is commonly assumed [8, 9].

For example, in the logistic model, let F(y) be the logistic cumulative distribution function,

$$F(y) = \frac{1}{1 + e^{-(\beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + X\beta)}} \qquad (8)$$

The interaction effect is then calculated as following according to the nature of the interaction:

- both X_1 and X_2 are continuous variables

$$\frac{\partial^2 F(y)}{\partial X_1 \partial X_2} = \beta_{12}\{F(y)(1 - F(y))\} + (\beta_1 + \beta_{12}X_2)(\beta_2 + \beta_{12}X_1) \times$$

$$[F(y)\{1 - F(y)\}\{1 - 2F(y)\}] \qquad (9)$$

- both X_1 and X_2 are dummy variables

$$\frac{\Delta^2 F(y)}{\Delta X_1 \Delta X_2} = \frac{1}{1 + e^{-(\beta_1+\beta_2+\beta_{12}+X\beta)}} - \frac{1}{1 + e^{-(\beta_1+X\beta)}} - \frac{1}{1 + e^{-(\beta_2+X\beta)}}$$

$$+ \frac{1}{1 + e^{-X\beta}} \qquad (10)$$

- X_1 is a continuous variable and X_2 is a dummy variable

$$\frac{\Delta \frac{\partial F(y)}{\partial X_1}}{\Delta X_2} = (\beta_1 + \beta_{12})(F\{(\beta_1 + \beta_{12})X_1 + \beta_2 + X\beta\} \times$$

$$(F\{(\beta_1 + \beta_{12})X_1 + \beta_2 + X\beta\})$$

$$-\beta_1[F(\beta_1 X_1 + X\beta)\{1 - F(\beta_1 X_1 + X\beta)\}] \qquad (11)$$

According to the above derivatives, it is clear that β_{12} alone is not a sufficient explanation when it comes to understanding the interaction effect on a dependent variable. The interaction effect does not have to be zero even when β_{12} is zero, since

$\frac{\partial^2 F(y)}{\partial X_1 \partial X_2}$ will then equal to $\beta_1\beta_2[F(y)\{1-F(y)\}\{1-2F(y)\}]$. The sign of interaction effect is also not necessary to be the same sign as of β_{12} for the same reason [8, 9].

2.2 Data Collection

Based on the literatures, the door holding gestures of the UT students may be accounted for the gender, distance between the ones opening the door and the recipients and the number of recipients.

The data was collected at three buildings with no automatic doors on the UT campus at randomly selected times. These buildings include, Robert Lee Moore Hall (RLM), Flawn Academic Center (FAC) and Gregory Gymnasium (GRE). For each observation, the following information was recorded:

- genders of the people opening the doors for others, and the recipients of this gesture. (The gender of the first recipients was recorded if there were multiple recipients);
- position of the person (whether it was the one who opened the door or the recipients);
- whether the person held the door for the recipients;
- distance between the one opening the doors and the recipients;
- the number of recipients.

The subjects were observed in an unobtrusive way. To obtain the distance between the people, tape of differing colors were used at every meter to mark different distances away from the door. The distance was then estimated according to the marks on the tape.

2.3 DATA ANALYSIS

Descriptive statistics is calculated in terms of the proportion of success as doors actually held under varying conditions.

For inferential statistics, logistic regression is performed to analyze the relationship between the response variable (conditions surrounding the door holding gesture) and a set of explanatory variables including gender, the position of the subject, the interaction term between gender and position, distance between the one opening the door and the recipients, and the number of recipients.

Microsoft Office Excel 2007 is used to complie descriptive statistics. The statistical software package STATA 9 is used to perform the logistic regression and analyze the relationship between the response variable and the explanatory variables. Predicted probabilities are plotted against different explanatory variables to visualize the relationships.

The interactions effect of gender and position in door holding is analyzed using both traditional method and Norton's method. The exponential of the coefficients associated with the interaction terms are commonly interpreted as the odds-ratio in the same way it is in linear models. A user written STATA command *inteff*, developed by Norton et al., is used to report the interaction effect on the response variable more correctly. This STATA post-estimation command is designed to run after fitting a logit or probit model [8, 9]:

Two graphs are produced to display the predicted probabilities of door holding conditions. The first plots interaction effects calculated by the conventional method (labeled as the "incorrect marginal effect") and by the method developed by Norton (labeled as the "correct interaction effect"). The second graph plots z-statistics of the interaction effect versus predicted probabilities, and is used to determine significance of interaction effect.

In summary, the logistic regression is well suited for the analysis of relationship between the response variable, door holding condition and a set of explanatory variables including gender, position, distance between people and the number of recipients. Both main and interaction effects will be studied. Especially, interaction effect will be analyzed using both the conventional method and the method developed by Norton to obtain the effect more correctly.

2.4 References

1. Cox, D. R., and Snell, E. J. 1989. *Analysis of Binary Data*. Second edition. London: Chapman & Hall.
2. Hosmer, D. W., and Lemeshow, S. 2000. *Applied Logistic Regression.* 2nd ed. John Wiley & Sons Inc.
3. Agresti, A. 2002. *Categorical Data Analysis.* New York: Wiley-Interscience.

4. Hilbe, J. M. 2009. *Logistic Regression Models.* Chapman & Hall/CRC Press

5. Balakrishnan, N. 1991. *Handbook of the Logistic Distribution.* Marcel Dekker, Inc..

6. Powers, A. P., and Xie, Y. 1999. *Statistical Methods for Categorical Data Analysis.* Academic Press, Inc.

7. Casella, G., and Berger, R. L. 2001. *Statistical Inference.* Second edition. Duxbury Press.

8. Ai, C., and Norton, E. C. 2003. Interaction Terms in Logit and Probit Models. *Economics Letters* 80:123–129.

9. Norton, E. C., Wang, H., and Ai, C. 2004. Computing Interaction Effects and Standard Errors in Logit and Probit Models. *The Stata Journal* 4(2): 154-167.

Chapter 3: Results and Conclusions

3.1 Descriptive statistics

3.3.1 Gender

Given that the person opening the door is a male, the probability that he will hold the door open for others ranges from 0.400 to 0.486, at RLM, FAC and GRE, with an average of 0.455. If the person opening the door is a female, the probability that she will hold the door open for others ranges from 0.450 to 0.500, with an average of 0.478 (Table 1 and Figure 2).

	RLM		FAC		GRE		Total	
	Male	Female	Male	Female	Male	Female	Male	Female
Held	68	27	56	30	66	30	190	87
	48.6%	45.0%	40.0%	50.0%	47.8%	48.4%	45.5%	47.8%
Not held	72	33	84	30	72	32	228	95
	51.4%	55.0%	60.0%	50.0%	52.2%	51.6%	54.5%	52.2%
Total	140	60	140	60	138	62	418	182

Table 1: Counts and percentages of door holding for others by gender of the person opening the door.

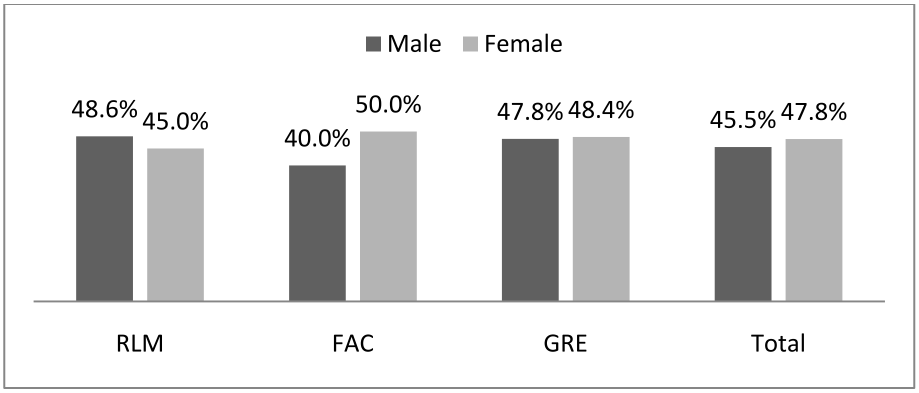

Figure 2: Door holding condition based on the gender of the person opening the door, reported in percentage of door held for recipients.

Table 2 and Figure 3 reveal the relationship between the door holding conditions and the gender of the recipient. If the recipient is a male, the probability that others will hold the door open for him ranges from 0.377 to 0.473, with an average of 0.438. Instead, if the recipient is a female, the probability that others will hold the door open for her ranges from 0.500 to 0.609, with an average of 0.534.

	RLM		FAC		GRE		Total	
	Male	Female	Male	Female	Male	Female	Male	Female
Held	72	23	58	28	69	27	199	78
	46.8%	50.0%	37.7%	60.9%	47.3%	50.0%	43.8%	53.4%
Not held	82	23	96	18	77	27	255	68
	53.2%	50.0%	62.3%	39.1%	52.7%	50.0%	56.2%	46.6%
Total	154	46	154	46	146	54	454	146

Table 2: Counts and percentages of door holding for others by gender of the recipients.

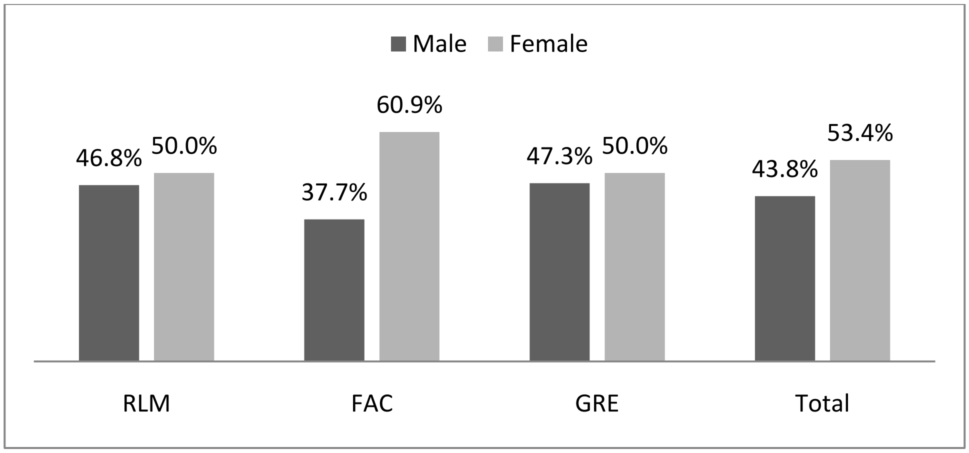

Figure 3: Door holding condition based on the gender of the recipients, reported in percentage of door held for recipients.

According to the above results, female recipients have a higher probability of having the door held open for them compared to their male counterparts. This pattern is consistent at all three locations. For the ones opening the doors, it is difficult without an inferential test result to determine whether women have a higher probability of holding the door open for others compared to men since the pattern is not consistent at all locations. For example, at RLM, men have a higher probability of holding the door open for others. In contrast, the data at other two locations yield opposite results. The significance of these differences requires further testing in the following logistic regression.

3.1.2 Distance

The door holding condition is also examined against distance between the one opening the doors and the recipients. According to the Figure 4, people are more reluctant to hold the door for the recipients when the distance between them increases, especially when the distance is more than three meters. The high percentages at four and four and a half meters are due to small number of observations at those large distances.

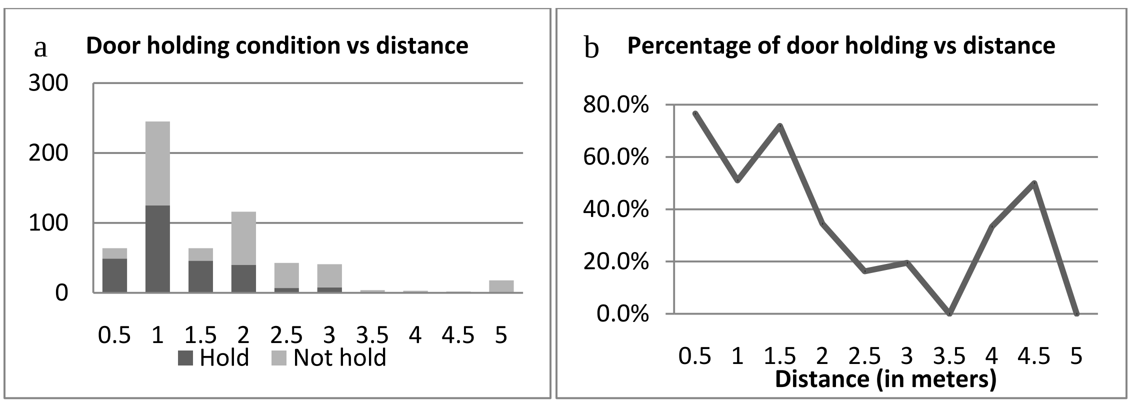

Figure 4: Door holding condition vs distance between the one opening the door and the recipients a) counts of door held and not held vs distance. b) door held in percentages vs distance.

3.1.3 Number of recipients

When examining the relationship between the door holding condition and the number of recipients, the analysis reveals that people have a higher probability of holding the door open if there are multiple recipients. Although the high percentages associated with four and five recipients may not be accurate due to small number of observations at those large numbers (Figure 5).

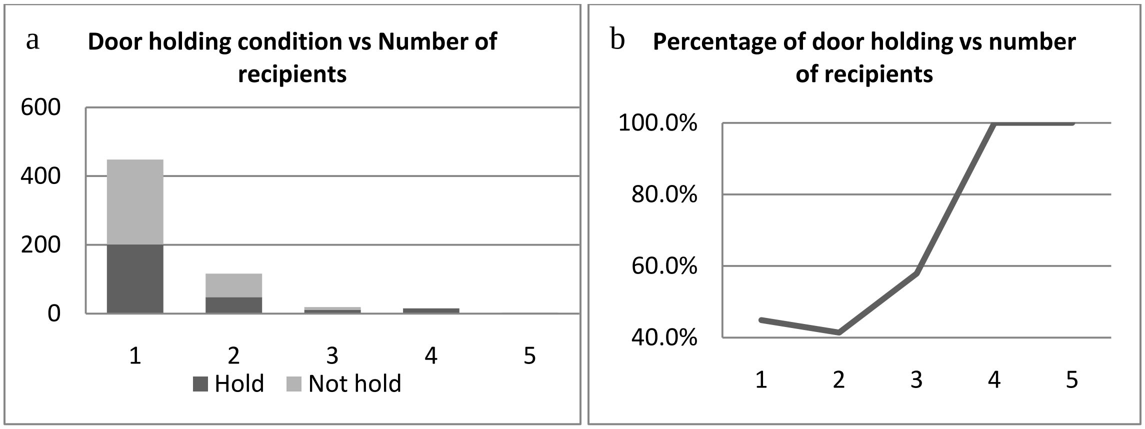

Figure 5: Door holding condition vs number of recipients. a) counts of door held and not held vs number of recipients. b) door held in percentages vs number of recipients.

3.2 Logistic regression

3.2.1 Model comparison

Logistic regressions are performed to further test the significance of the above observations. The proposed logit models are initially:

Model 1:

$$\text{Logit(p)} = \ln\left(\frac{\text{p}}{1-\text{p}}\right) = \beta_0 + \beta_1 * gender + \beta_2 * front + +\beta_{12} * Interact_{Gender\times front}$$

$$+\beta_3 * distance + \beta_4 * \#\, recipients + \beta_5 * I_{RLM} + \beta_6 * I_{FAC} \qquad (16)$$

Model 2:

$$\text{Logit(p)} = \ln\left(\frac{\text{p}}{1-\text{p}}\right) = \beta_0 + \beta_1 * gender + \beta_2 * front + +\beta_{12} * Interact_{Gender \times front}$$

$$+\beta_3 * distance + \beta_4 * \#\, recipients \qquad (17)$$

Compared to model 1, model 2 is simpler since it does not consider the location difference as a factor. Both models are analyzed and the results are summarized in STATA output below (Tables 3 and 4). STATA user written command *fitstat* is used for model comparison and selection (Table 5).

```
. logistic holding gender front genfront distance nofollow l1 l2

Logistic regression                               Number of obs   =       1200
                                                  LR chi2(7)      =     171.03
                                                  Prob > chi2     =     0.0000
Log likelihood = -742.73316                       Pseudo R2       =     0.1032
```

holding	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
gender	.5535821	.1142389	-2.87	0.004	.3694243	.8295425
front	.600672	.1422913	-2.15	0.031	.3775712	.9555995
genfront	1.982633	.5553154	2.44	0.015	1.145063	3.432853
distance	.4038697	.0352863	-10.38	0.000	.3403073	.4793042
nofollow	1.305847	.1217801	2.86	0.004	1.087705	1.567737
l1	.9705619	.1480475	-0.20	0.845	.7197517	1.308771
l2	.8127674	.1240616	-1.36	0.174	.6026124	1.096212

Table 3: STATA logistic regression output for model 1. L1 and l2 are location dummy variables, where l1=1 if the building is RLM and 0 otherwise; l2=1 if the building is FAC and 0 otherwise.

```
. logistic holding gender front genfront distance nofollow

Logistic regression                               Number of obs   =       1200
                                                  LR chi2(5)      =     168.86
                                                  Prob > chi2     =     0.0000
Log likelihood = -743.81824                       Pseudo R2       =     0.1019
```

holding	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
gender	.551538	.1137303	-2.89	0.004	.3681738	.8262245
front	.6000486	.1420638	-2.16	0.031	.3772778	.9543587
genfront	1.984271	.5553747	2.45	0.014	1.146462	3.434334
distance	.4045779	.035323	-10.36	0.000	.3409458	.4800859
nofollow	1.316239	.1221356	2.96	0.003	1.097363	1.57877

Table 4: STATA logistic regression output for model 2.

```
. fitstat, using(m0)

Measures of Fit for logistic of holding
```

	Current	Saved	Difference
Model:	logistic	logistic	
N:	1200	1200	0
Log-Lik Intercept Only:	-828.246	-828.246	0.000
Log-Lik Full Model:	-743.818	-742.733	-1.085
D:	1487.636(1194)	1485.466(1192)	2.170(2)
LR:	168.856(5)	171.027(7)	-2.170(-2)
Prob > LR:	0.000	0.000	-0.000
McFadden's R2:	0.102	0.103	-0.001
McFadden's Adj R2:	0.095	0.094	0.001
Maximum Likelihood R2:	0.131	0.133	-0.002
Cragg & Uhler's R2:	0.175	0.177	-0.002
McKelvey and Zavoina's R2:	0.206	0.208	-0.002
Efron's R2:	0.128	0.129	-0.001
Variance of y*:	4.144	4.152	-0.009
Variance of error:	3.290	3.290	0.000
Count R2:	0.630	0.630	0.000
Adj Count R2:	0.199	0.199	0.000
AIC:	1.250	1.251	-0.002
AIC*n:	1499.636	1501.466	-1.830
BIC:	-6977.915	-6965.905	-12.010
BIC':	-133.406	-121.396	-12.010

```
Difference of   12.010 in BIC' provides very strong support for current model.
```

Note: Saved: Model 1; Current: Model 2.

Table 5: STATA fitstat output: models goodness-of-fit comparison.

The results of goodness-of-fit comparison between model 1 and model 2 show that the Bayesian information criterions (BIC') are -121.396 for model 1 and -133.406 for model 2. The difference of 12.01 in BIC' provides very strong evidence that model 2 is a better model compared to model 1 (Table 5).

Additionally, the z-statistics of the two location indicators in model 1 are -0.2 and -1.36, and the associated p-values are 0.845 and 0.174, respectively. At the default significance level of α=0.05, we can conclude that the location does not have a statistically significant effect on door holding behavior. This conclusion also suggests that removing these two location indicators from the model would prove beneficial according to the rule of parsimony/simplicity in model selection [1]. Therefore, model 2 is selected for further analysis.

3.2.2 Goodness-of-fit and interpretation of the coefficients

For the selected model, the log likelihood for intercept alone is -828.246 (L_0), and the log likelihood for the full model is -743.818 (L_1) (Table 5). The likelihood ratio chi-square with 5 degrees of freedom is then calculated as $2*(L_1 - L_0)$ =168.856. The p-value associated with this chi-square value is very low. The nearly zero p-value indicates the model as whole was statistically significant.

The pseudo R-squared statistic vary from 0.095 to 0.206 through the use of different calculation methods. Although the pseudo R-squared statistic in the logistic regression cannot be interpreted in the same way as its counterpart in linear regression, researchers used simulations to predict a continuous, latent variable through the ordinary least square (OLS) regression and its observed binary variable through logistic regression, and compare the pseudo R-squared to the OLS R-squared. The result showed that pseudo R-squared calculated by McKelvey and Zavoina method was the closest to the OLS R-squared [2, 3]. The McKelvey and Zavoina pseudo R-squared in this analysis is 0.206. Since it is the closest to the OLS R-squared, it suggests that the variability in door holding condition is accounted for the set of explanatory variables (gender, position, distance, number of recipients and interaction term between gender and position) by 20.6%, which is moderately high for cross-sectional data in the social sciences.

In Table 4, the p-values associated with main factors and interaction term suggest that all main and interaction effects are statistically significant at the significance level of α=0.05. The effects of individual explanatory variables on door holding behavior are explained below in more detail.

- Gender and position

  The odds ratios are 0.55 for gender and 1.98 for the interaction term between gender and position. When focused on the one opening the door, the odds that men will hold the door for the recipients are a multiplicative factor of 1.09 (=0.55*1.98) of that for women on average, holding all other factors constant; However, women have much higher probability of having doors held open for them. The odds for men are a multiplicative factor of 0.55 of that for women on average, holding other factors constant.

- Distance

  The p-value associated with distance is close to zero, providing very strong evidence that distance has a significant relationship with the door holding condition. The odds ratio for distance is 0.40. This result suggests that for every meter increase in distance between the ones who open the door and the recipients, the odds will decrease by a multiplicative factor of 0.40, on average, holding all other factors constant. We are also 95% confident that the odds will decrease multiplicatively between 0.34 and 0.48 for every meter increase in distance. The illustrated graph of predicted probabilities of door held against distance shows that the predicted probability decreases quickly when the distance increases. Men have a slightly higher probability of holding the door for others than women (Figure 6); while women have a much higher probability of having someone hold the door for them (Figure 7). The number of recipients is presumably one in calculation for purposes of simplicity.

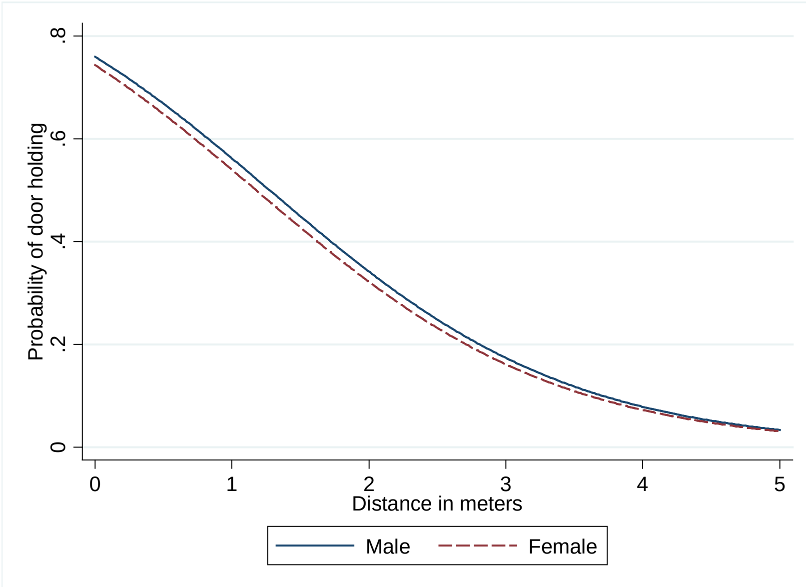

Figure 6: Predicted probability of a person holding the door for others against distance.

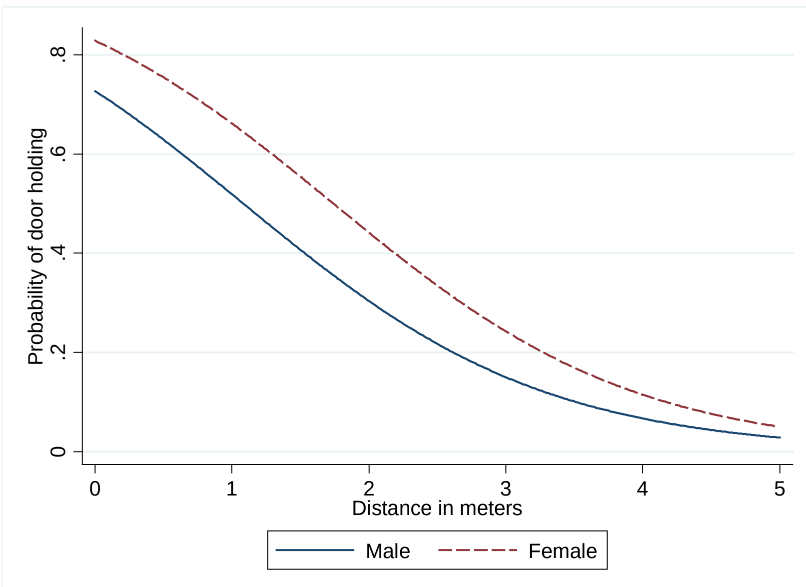

Figure 7: Predicted probability of others holding the door for a person against distance.

- Number of recipients

  The p-value associated with distance is only 0.003, providing very strong evidence that distance had a significant relationship with door-holding condition.

  The odds ratio for distance is 1.32. This ratio suggests that for every increase in the number of recipients, on average, the odds will increase by a multiplicative factor of 1.32, holding all other factors constant. We are also 95% confident that the odds will increase multiplicatively between 1.10 and 1.58 for every increase in number of recipients. The graph of predicted probability of doors held against the number of recipients reveals that the probability increases roughly linearly as the number of recipients increases. Men have a higher probability of holding the door for others than women (Figure 8); while women have a much higher probability of having others hold the door for them (Figure 9). The distance between these people is presumably one meter in calculation for purposes of simplicity.

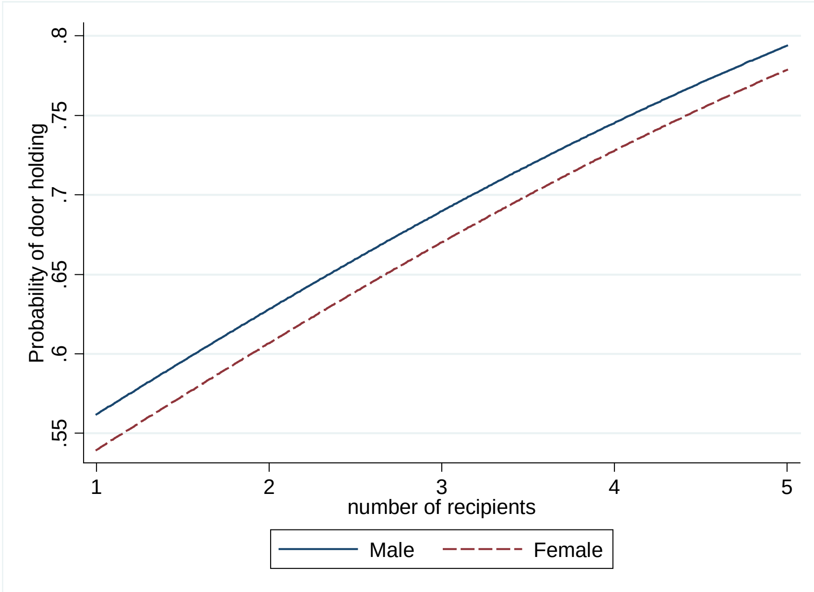

Figure 8: Predicted probability of people holding the door for others against the number of recipients.

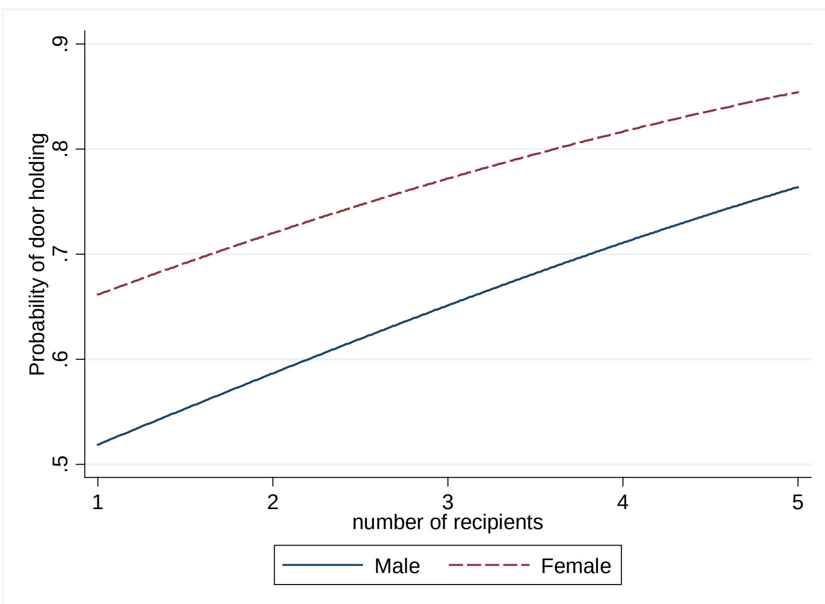

Figure 9: Predicted probability of people having the door held by others against the number of recipients.

3.2.3 The interaction effect of gender with position

In the conventional interpretation, the interaction effect between gender and position is estimated here to have marginal odds of 0.69 on average with a standard error of 0.28. The effect is statistically significant. However, as noted on pp. 5-7, the correct interpretation of interaction terms in logit/probit models is more problematic than is commonly recognized. In order to get a more accurate result on interaction effect, Norton's user written command *inteff* is used [4]. The result is shown in Table 6.

Variable	Obs	Mean	Std. Dev.	Min	Max
ie	1200	.1482024	.0305196	.0244604	.1700827
se	1200	.0602957	.0117344	.0119908	.0688495
z	1200	2.443016	.0848102	2.039932	2.48772

Table 6: STATA output summary for command *inteff*

In the above results, variable ie is a measure of interaction effect in percentage points and z is the corresponding z-statistics. The mean interaction effect is 0.15 with a standard deviation of 0.03. Although the results are not the same as the ones obtained using the traditional method, the distributions of the corrected interaction effect and the incorrect marginal effect are similar, and both of their interact effect percentages are above zero (Figure 10). The z-statistics of the interaction effect are also plotted in Figure 11. The mean z-statistic is 2.44, with a range between 2.04 and 2.49. These z-statistics suggest the interaction effect is significant at the default significance level of 0.05.

Therefore, both the conventional and corrected interpretation of the interaction effect of gender and position reach the same conclusion on interaction effect in this study: interaction between gender and position has a statistically significant effect on door holding conditions.

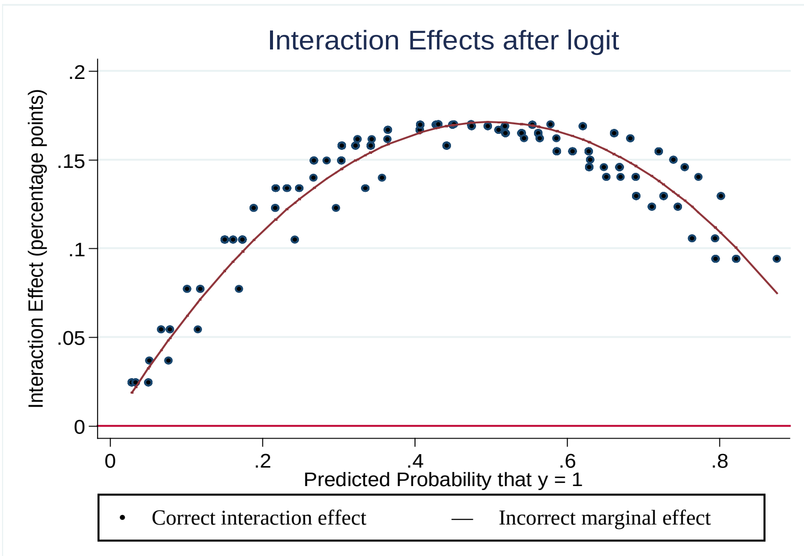

Figure 10: Interaction effect and marginal effect against predicted probability.

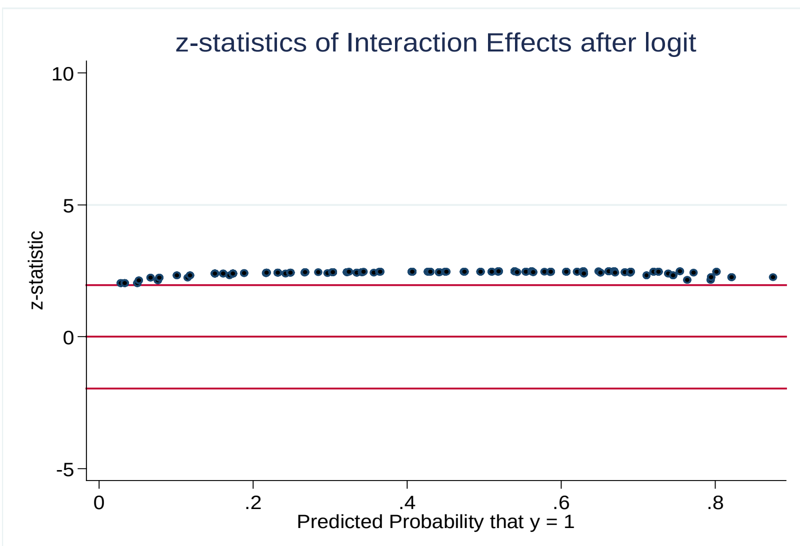

Figure 11: Z-statistics of interaction effect against predicted probability.

3.3 CONCLUSION

In conclusion, door-holding behavior at the University of Texas at Austin are significantly affected by gender, position, distance between the ones opening the doors and the recipients of these courtesies, the number of recipients and the interaction of gender and position. Men have a slightly higher probability of holding the door open for the recipients. On the other hand, people are more likely to hold the door open for women. Gender still plays an important role in this simple social courtesy.

Distance between people and the number of recipients are the other two factors affect this gesture. The probability of doors held decrease when the distance becomes longer or when there are fewer recipients.

This study was carried out in an unobtrusive manner without disturbing the subjects. Compared to survey studies, more honest results are expected and the findings more convincing. However, there are still problems. For instance, we did not know if the pair of people were acquainted with each other. The degree of acquaintance very well affect door-holding behaviors. Other factors such as weather conditions and time may also play a role in door holding conditions. Incoporating these factors into the study may improve the model and provide more accurate predictions and greater external varidity.

3.4 References

1. Burnham, K. P., and Anderson, D. R. 2002. *Model Selection and Multimodel Inference: A Practical-Theoretic Approach*, 2nd ed. New York: Springer-Verlag.
2. Freese, J., and Long, J. S. 2006. *Regression Models for Categorical Dependent Variables Using Stata.* College Station: Stata Press.
3. Long, J. S. 1997. *Regression Models for Categorical and Limited Dependent Variables.* Thousand Oaks: Sage Publications.
4. Norton, E. C., Wang, H., and Ai, C. 2004. Computing Interaction Effects and Standard Errors in Logit and Probit Models. *The Stata Journal* 4(2): 154-167.

Appendix

STATA Syntax

Sample predicted probability graphing (thanks to Dr. Stolp for template)

```
twoway (function
y=invlogit(_b[gender]+_b[front]+_b[genfront]+_b[distance]*x+_b[nofollow]+_b[_cons])
, range(0 5))     ///
(function y=invlogit(_b[front]+_b[distance]*x+_b[nofollow]+_b[_cons]), range(0 5)
clpatt(dash)),      ///
xtitle("Distance") ytitle("Probability of door held") ///
legend(order(1 "Male" 2 "Female")) xline(1.00) xline(4.00)
```

User written STATA commands

fitstat

Long, J. S. and Freese, J. 2001. FITSTAT: Stata module to compute fit statistics for single equation regression models. Boston College: Research Papers in Economics. http://fmwww.bc.edu/repec/bocode/f/fitstat.ado

inteff

Norton, E. C., Wang, H., and Ai, C. 2004. Computing Interaction Effects and Standard Errors in Logit and Probit Models. http://www.stata-journal.com/software/sj4-2/st0063/

Bibliography

Agresti, A. 2002. *Categorical Data Analysis.* New York: Wiley-Interscience.

Ai, C., and Norton, E. C. 2003. Interaction Terms in Logit and Probit Models. *Economics Letters* 80:123–129.

Balakrishnan, N. 1991. *Handbook of the Logistic Distribution.* Marcel Dekker, Inc..

Burnham, K. P., and Anderson, D. R. 2002. *Model Selection and Multimodel Inference: A Practical-Theoretic Approach*, 2nd ed. New York: Springer-Verlag.

Casella, G., and Berger, R. L. 2001. *Statistical Inference.* Second edition. Duxbury Press.

Cox, D. R., and Snell, E. J. 1989. *Analysis of Binary Data*. Second edition. London: Chapman & Hall.

Freese, J., and Long, J. S. 2006. *Regression Models for Categorical Dependent Variables Using Stata.* College Station: Stata Press.

Gibson, E. K. 2009. Gender, Polite Questions and the Fast-food Industry. *Griffith Working Papers in Pragmatics and Intercultural Communication* 2(1):1-17.

Hilbe, J. M. 2009. *Logistic Regression Models.* Chapman & Hall/CRC Press.

Hosmer, D. W., and Lemeshow, S. 2000. *Applied Logistic Regression.* 2nd ed. John Wiley & Sons Inc.

Lakoff, R. T. 2004. *Language and Women's Place.* Rev. and expanded ed. New York; Oxford: Oxford University Press.

Long, J. S. 1997. *Regression Models for Categorical and Limited Dependent Variables.* Thousand Oaks: Sage Publications.

Montgomery, M. B. 1998. "Multiple Modals in LAGS and LAMSAS," In *From the Gulf States and Beyond: The Legacy of Lee Pederson and LAGS,* edited by Nunnally, T., 90-122, Tuscaloosa/London: University of Alabama Press.

Norton, E. C., Wang, H., and Ai, C. 2004. Computing Interaction Effects and Standard Errors in Logit and Probit Models. *The Stata Journal* 4(2): 154-167.

Powers, A. P., and Xie, Y. 1999. *Statistical Methods for Categorical Data Analysis.* Academic Press, Inc..

Wouters, C. 2004. *Sex and Manners: Female Emancipation in the West 1890-2000.* London: SAGE Publications.