

# A quantile regression of personal network size

Itaru Ishiguro

Japan Women's University

Stata User Meeting

2013.9.20

# Attention Limited to Means

- Sociological and psychological researchers examine only the central positions of target variables (i.e., mathematical means).
- However, there are no theoretical reasons to limit their attention to these central positions.
- The only reason is that the targets of standard statistical analysis are usually mathematical means.
  - Also, they do not know how to treat a property of the variables other than through means.
- Quantile regression changes this situation.

# Why is the Limitation Problematic?

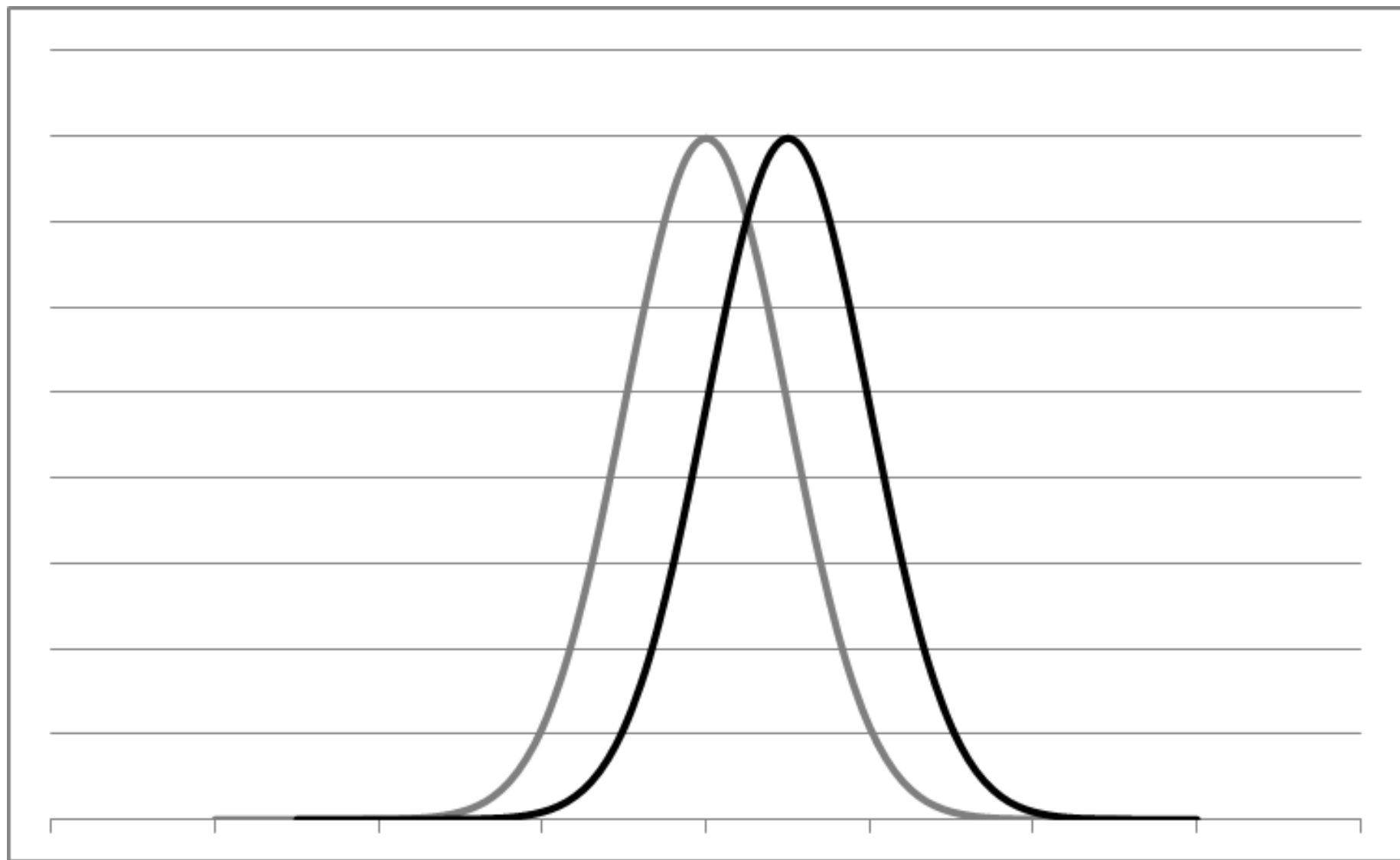
(1) Means are vulnerable to outliers.

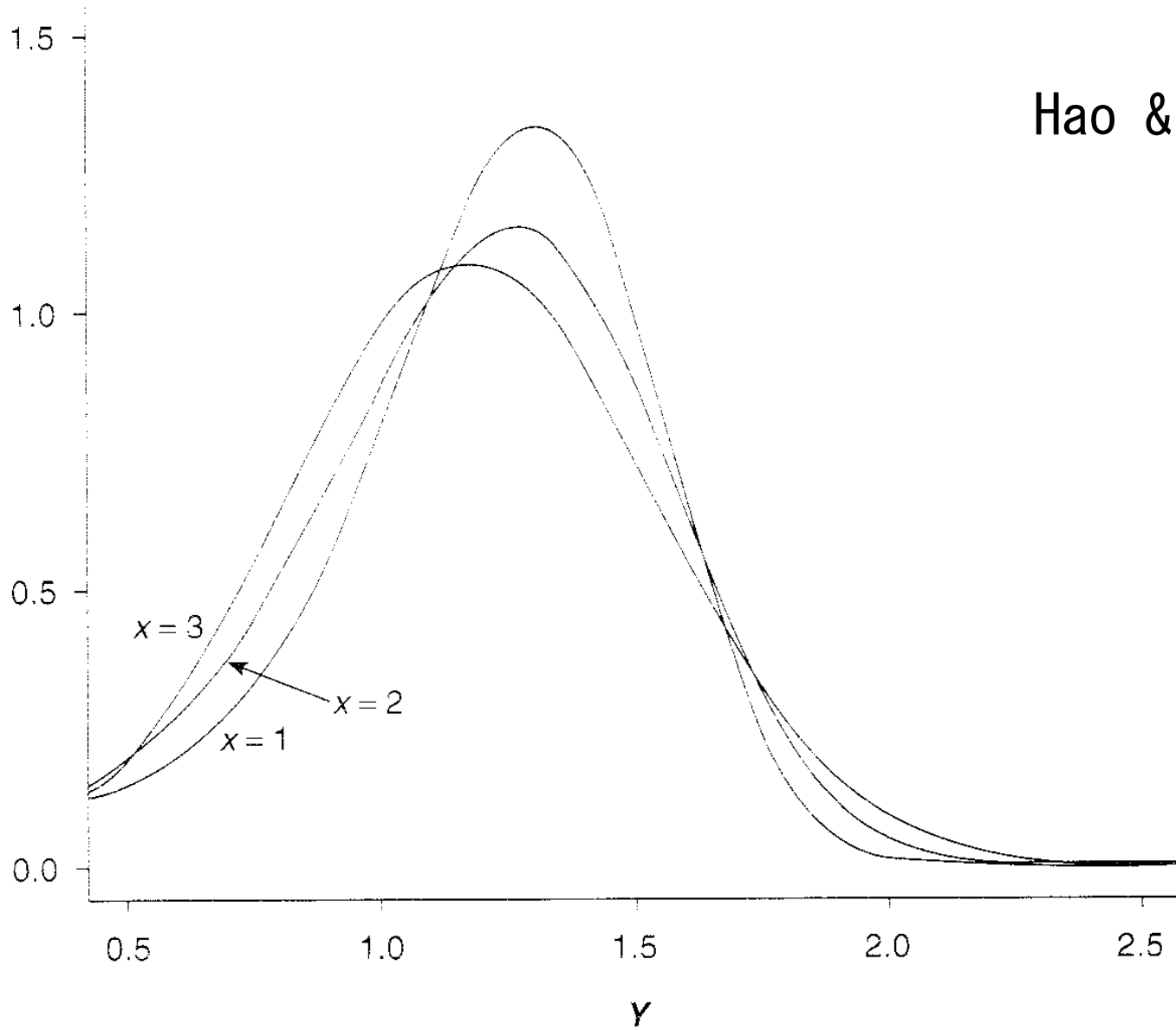
...However, outliers are essential in some research fields (e.g., the size of personal networks).

(2) Properties other than the mean are ignored.

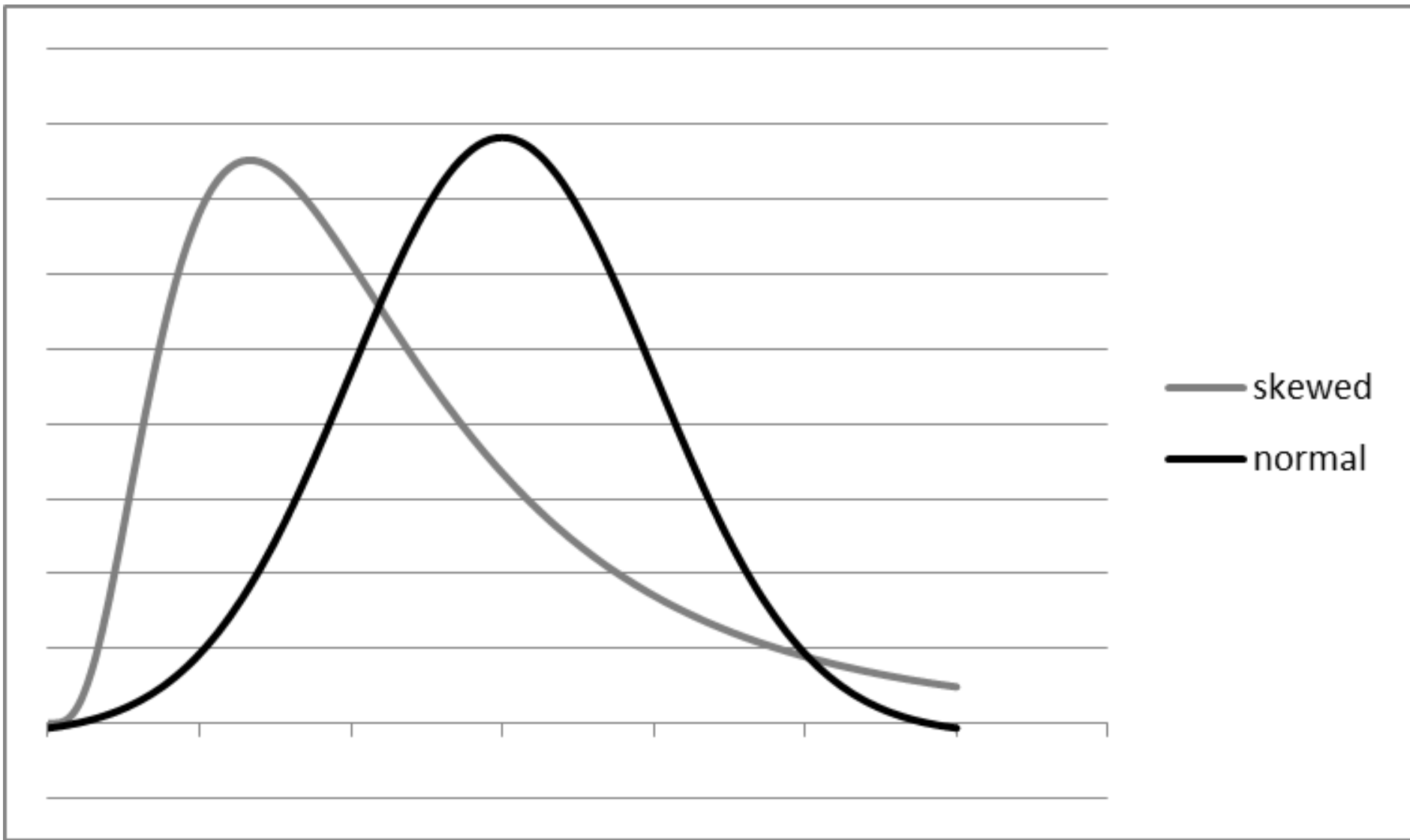
...The assumption of homoscedasticity and normality is very strong.

...We cannot always detect changes in the shape of distribution by examining means and standard deviations.





**Figure 3.1** Conditional Distributions With the Same Mean and Standard Deviation but Different Skewness



Conditional distributions with the same means and slightly different standard deviations

# Why The Limitation is problematic?

(3) Outliers/Shapes of the distribution should be the target of research in some areas.

...E.g., in studies of

...Social Inequality.

...Well-being.

...Network Size.

# “Difficulties” in Measuring and Analyzing Network Size

- The size of egocentric networks (network size) captures the attention of researchers of sociology, psychology, and anthropology.
- However, the measurement and analysis of network size is difficult.
  - Network size is right-skewed
  - A large percentage of respondents answer that their size of their personal network is zero.



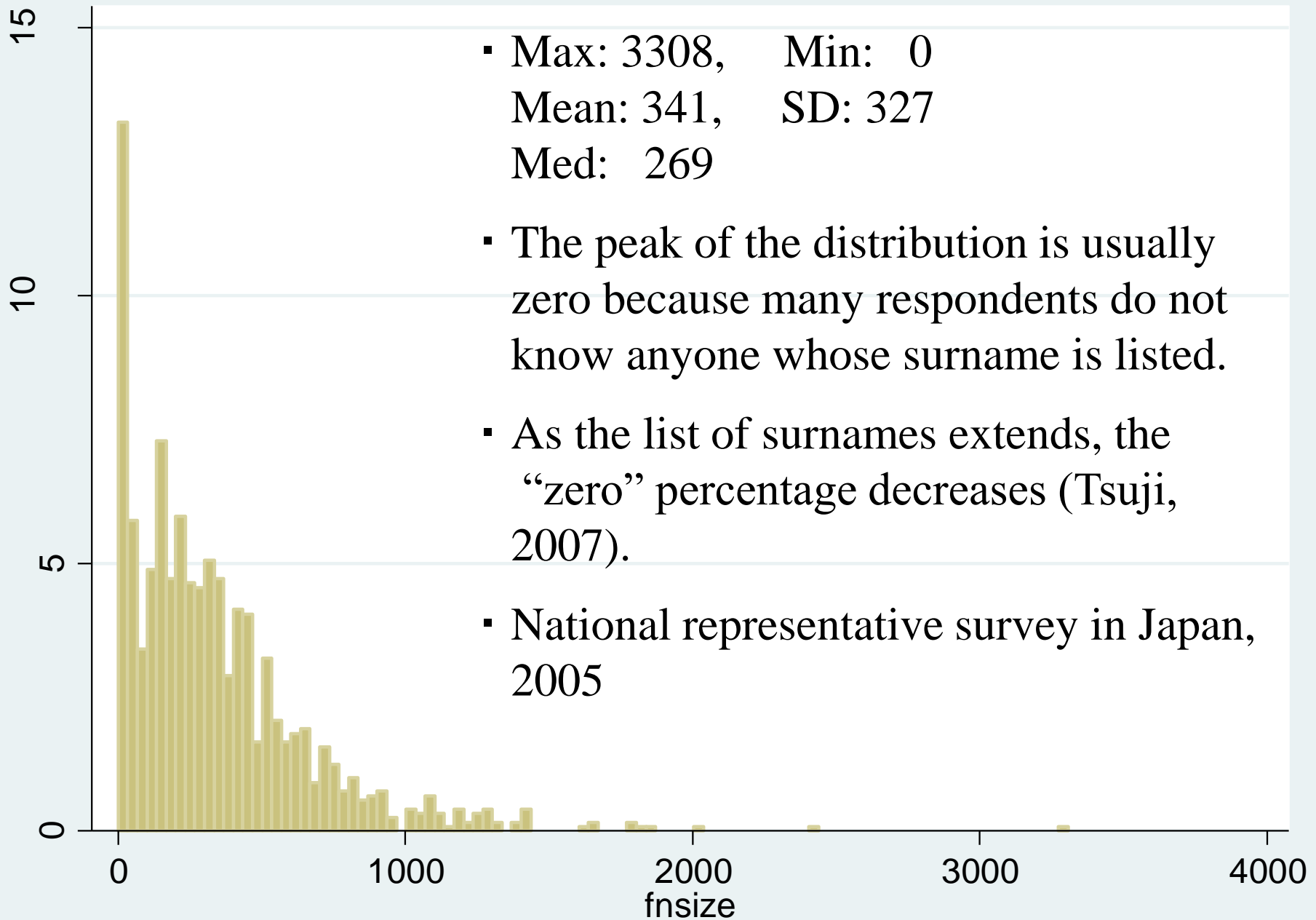
# “Address book method” to measure “acquaintance volume”

次にいくつかの苗字（みょうじ）をお見せします。あなたのお知り合いの中で、次のような苗字の方は何人いますか。それぞれ人数をお書きください。そのような苗字のお知り合いがない場合は、それぞれ（ 〇 ）人とお書きください。

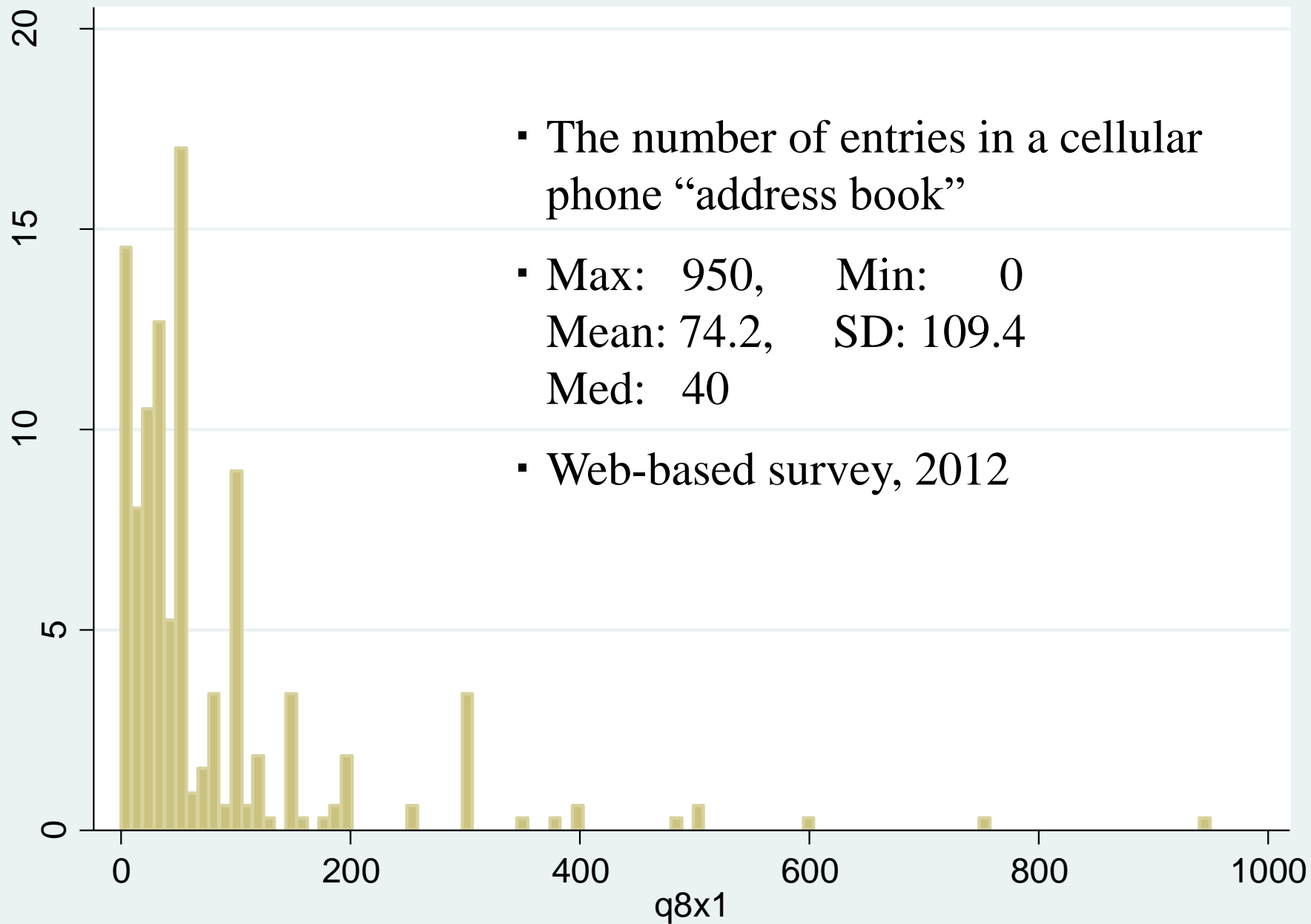
(ア) 岩崎	(            ) 人	(カ) 和田	(            ) 人
(イ) 平野	(            ) 人	(キ) 田村	(            ) 人
(ウ) 上野	(            ) 人	(ク) 中野	(            ) 人
(エ) 高田	(            ) 人	(ケ) 松田	(            ) 人
(オ) 石田	(            ) 人	(コ) 青木	(            ) 人

上にあげた、どの苗字のお知り合いもない場合は、次の（    ）の中に〇をつけてください……………（            ）

- Developed by Freeman & Thompson(1989) and Killworth et al. (1990), and Tsuji (2007) for Japanese.
- Summing the answers with the weights based on the distribution of each name among the population.



- Max: 3308, Min: 0  
Mean: 341, SD: 327  
Med: 269
- The peak of the distribution is usually zero because many respondents do not know anyone whose surname is listed.
- As the list of surnames extends, the “zero” percentage decreases (Tsuji, 2007).
- National representative survey in Japan, 2005



- The number of entries in a cellular phone “address book”
- Max: 950, Min: 0  
Mean: 74.2, SD: 109.4  
Med: 40
- Web-based survey, 2012

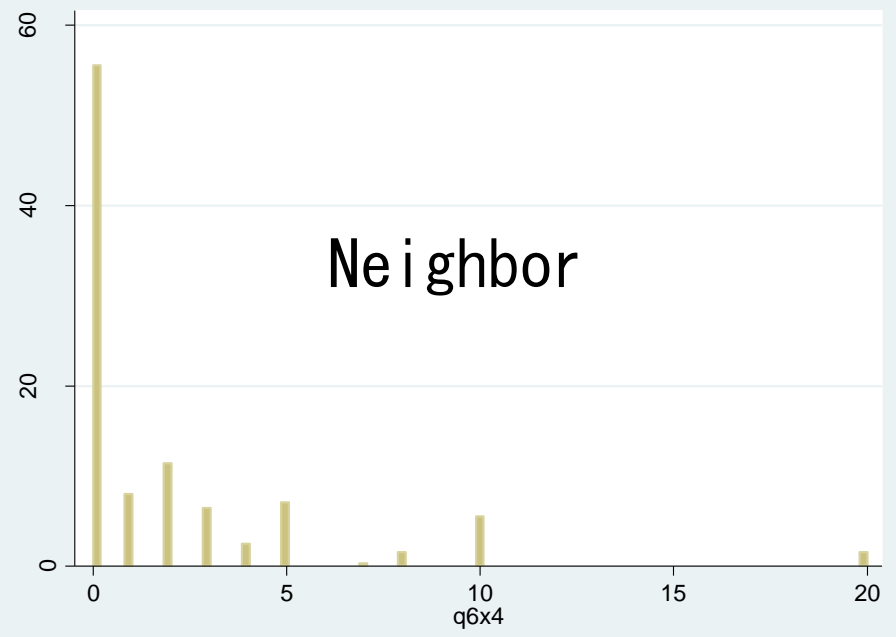
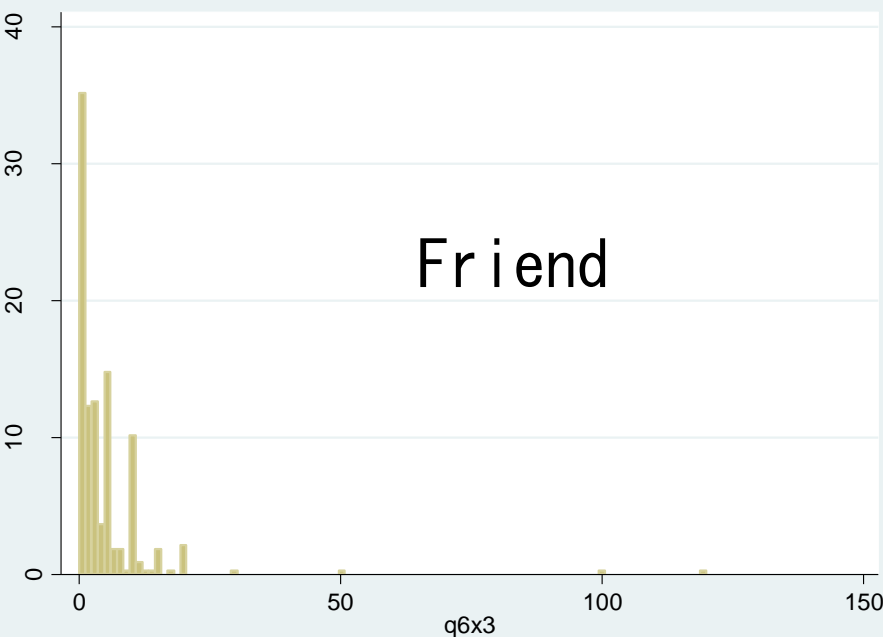
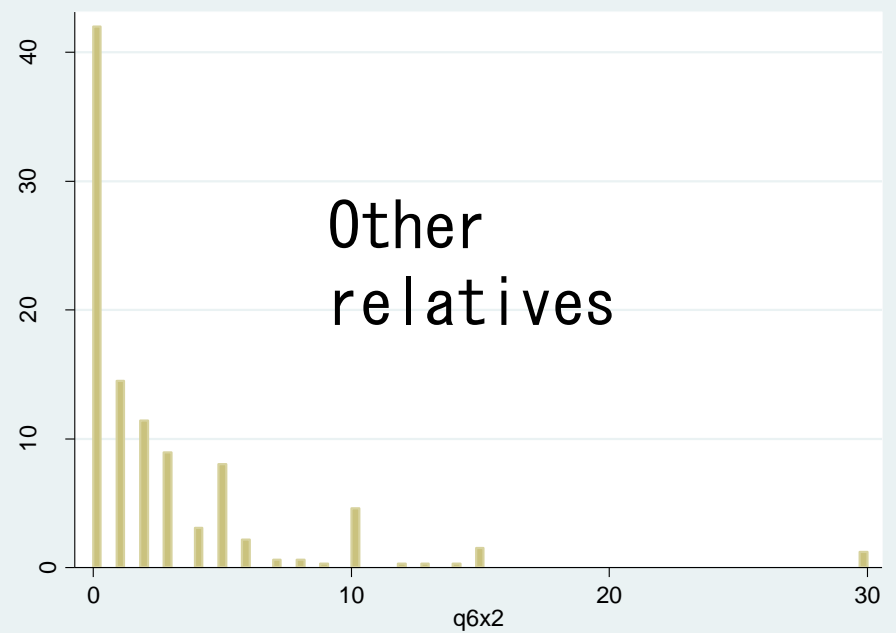
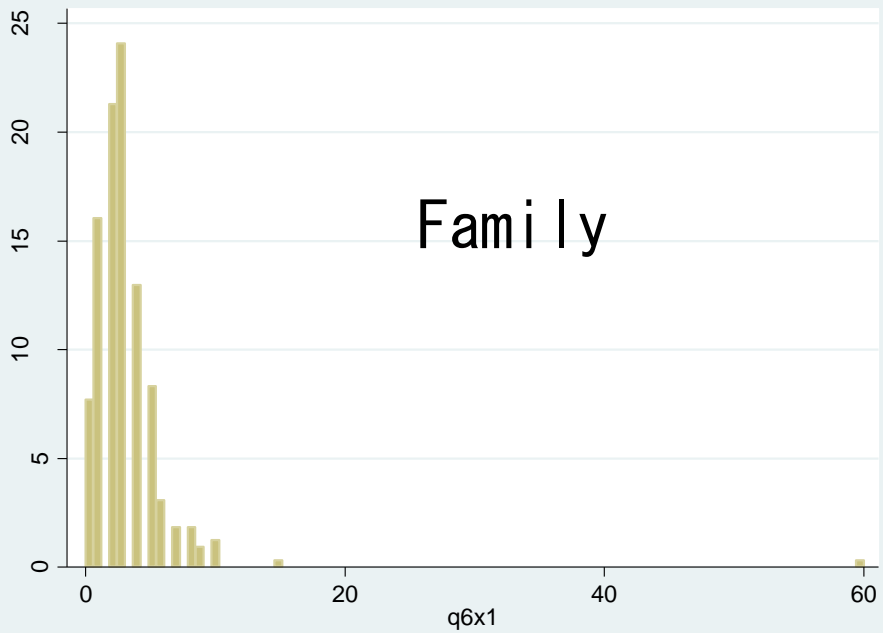
# Simple Measurement Method of the Sympathy Group (Stiller & Dunbar, 1995)

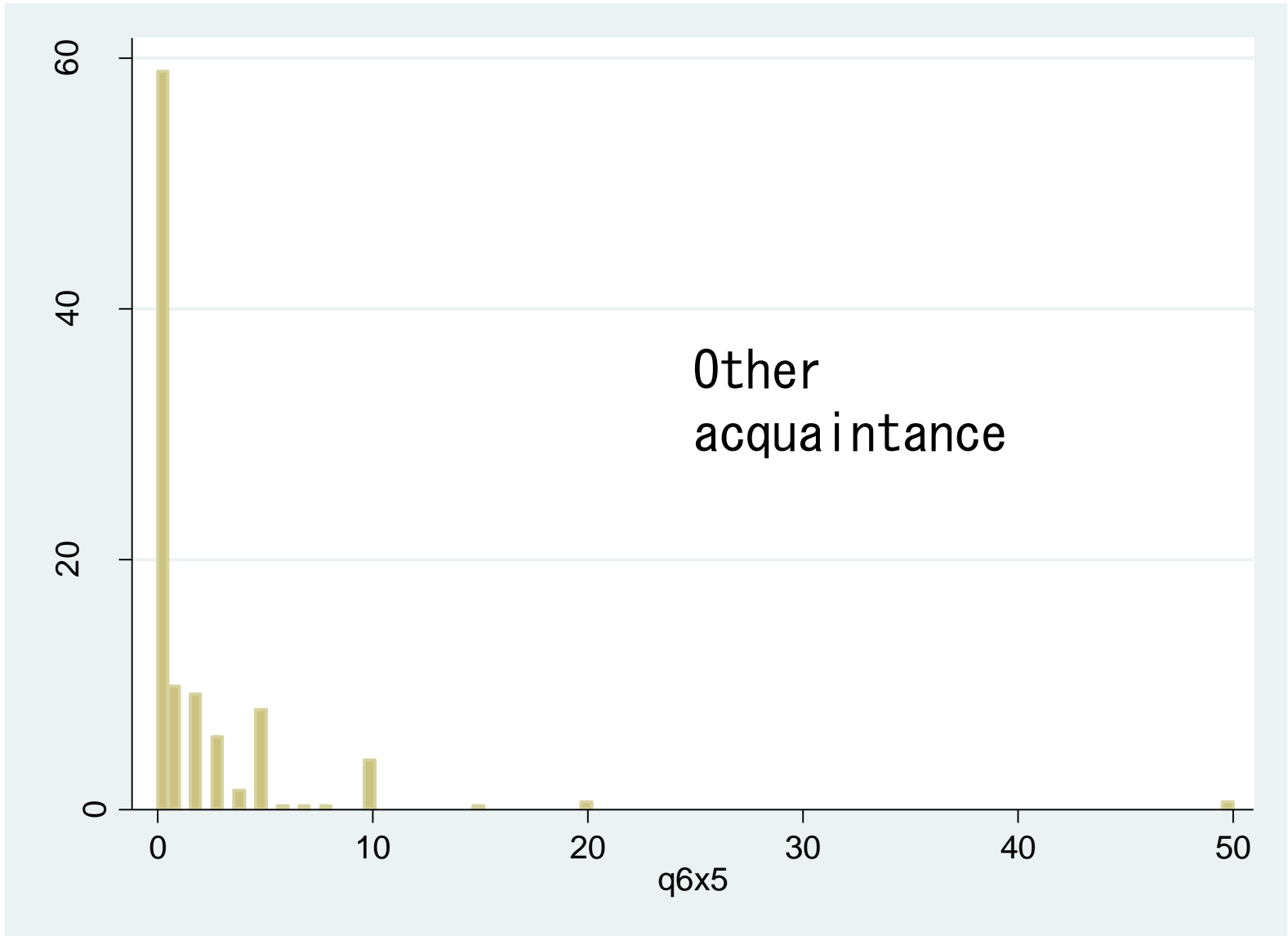
**Q6. 過去一ヶ月の間に、仕事以外の用件で会うか連絡した家族、親戚、友人、ご近所、それ以外の知人は何人くらいいますか。いない場合は「0」とお答えください。**

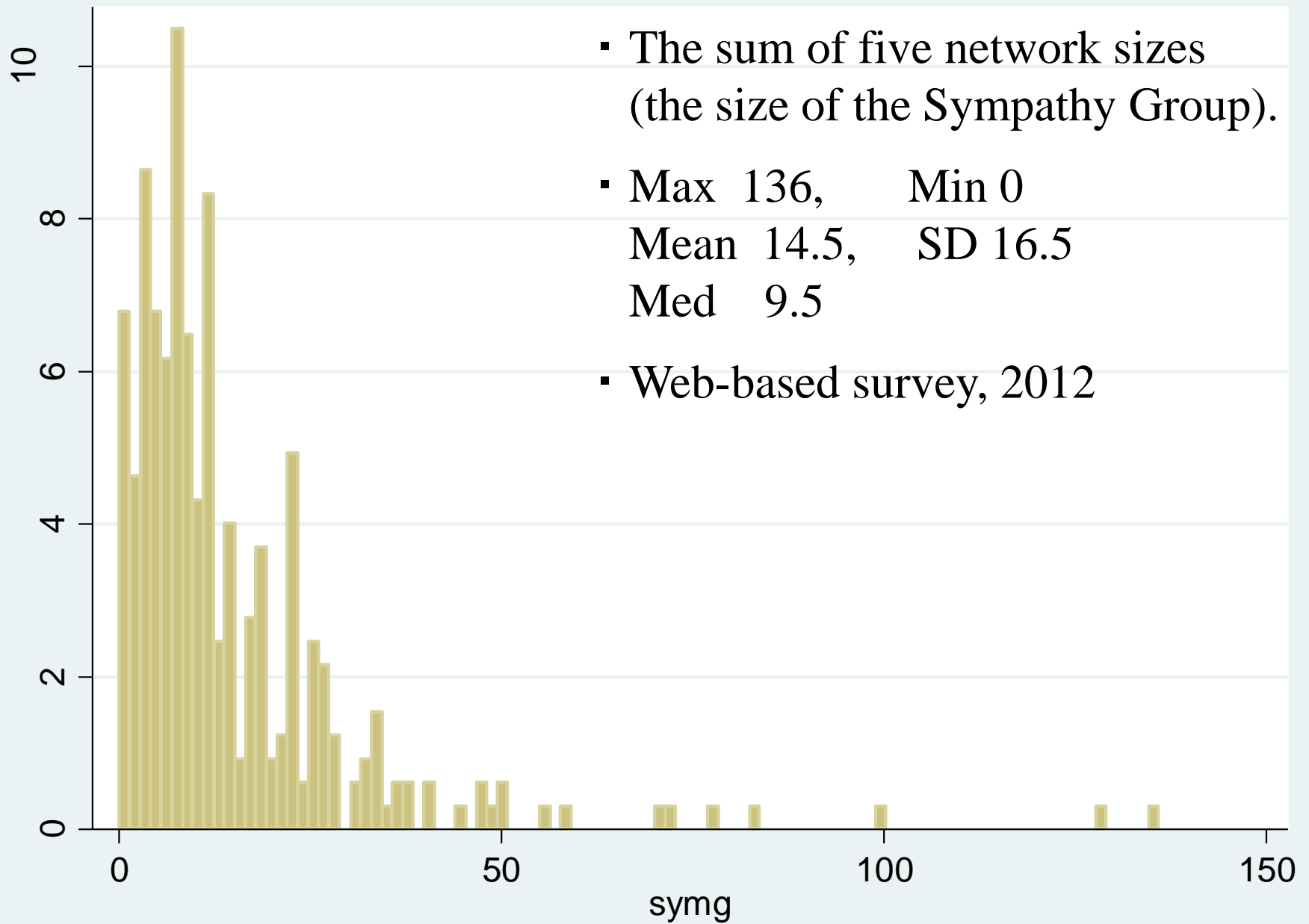
---

回答は、半角数字でお願いいたします。

家族	<input type="text"/>	人くらい
親戚	<input type="text"/>	人くらい
友人	<input type="text"/>	人くらい
ご近所	<input type="text"/>	人くらい
それ以外の知人	<input type="text"/>	人くらい







# Outliers?

- Respondents with extremely large networks are not merely “outliers.” They are network hubs (Barabasi, 2002) .
- Some researchers are interested in network hubs, and they cannot drop them as outliers.
  - Rather, their exploration is based on the variables with which they can predict the emergence of network hubs.
- We need a method
  - that is not vulnerable to outliers, and
  - That can analyze the shape of the distribution itself.



# Quantile Regression

Koenket & Bassett(1978), Hao & Naiman(2007)

- A method to predict any multiple quantiles (percentiles) with linear equations like standard regression.

$$y_i^{(p)} = \sum_i b_i^{(p)} x_i^{(p)} + e_i^{(p)}$$

- Parameters are estimated by minimizing weighted mean deviation. SEs are estimated by bootstrap.
- Economic researchers have widely used quantile regression since 00s, though the method was established in 70s.

# The benefits of quantile regression

- (1) Can analyze a change of shape of the distribution of target variables.  
...by analyzing multiple quantiles simultaneously and comparing the slopes of explanatory variables.
  
- (2) Can analyze a change of the central position of the distribution while minimizing the effect of outliers.  
...because medians are less vulnerable to outliers than means.

# Stata Commands

- To predict single quantile

```
bsqreg depvar [indepvars] [if] [in]  
      [, bsqreg_options]
```

- To predict multiple quantiles simultaneously

```
sqreg depvar [indepvars] [if] [in]  
      ,q(# # #...) [ sqreg_options]
```

# Example

Target variable: sympathy group size

Target quantiles: From .10th to .90th quantiles (intervals of 10)

Explanatory variable: extraversion(centered), gender, age(centered), and education years(centered)

```
set seed #
```

```
sqreg symg cext cage sex cedu,  
q(10 20 30 40 50 60 70 80 90) rep(500)
```

# Quantile regression of the sympathy group size

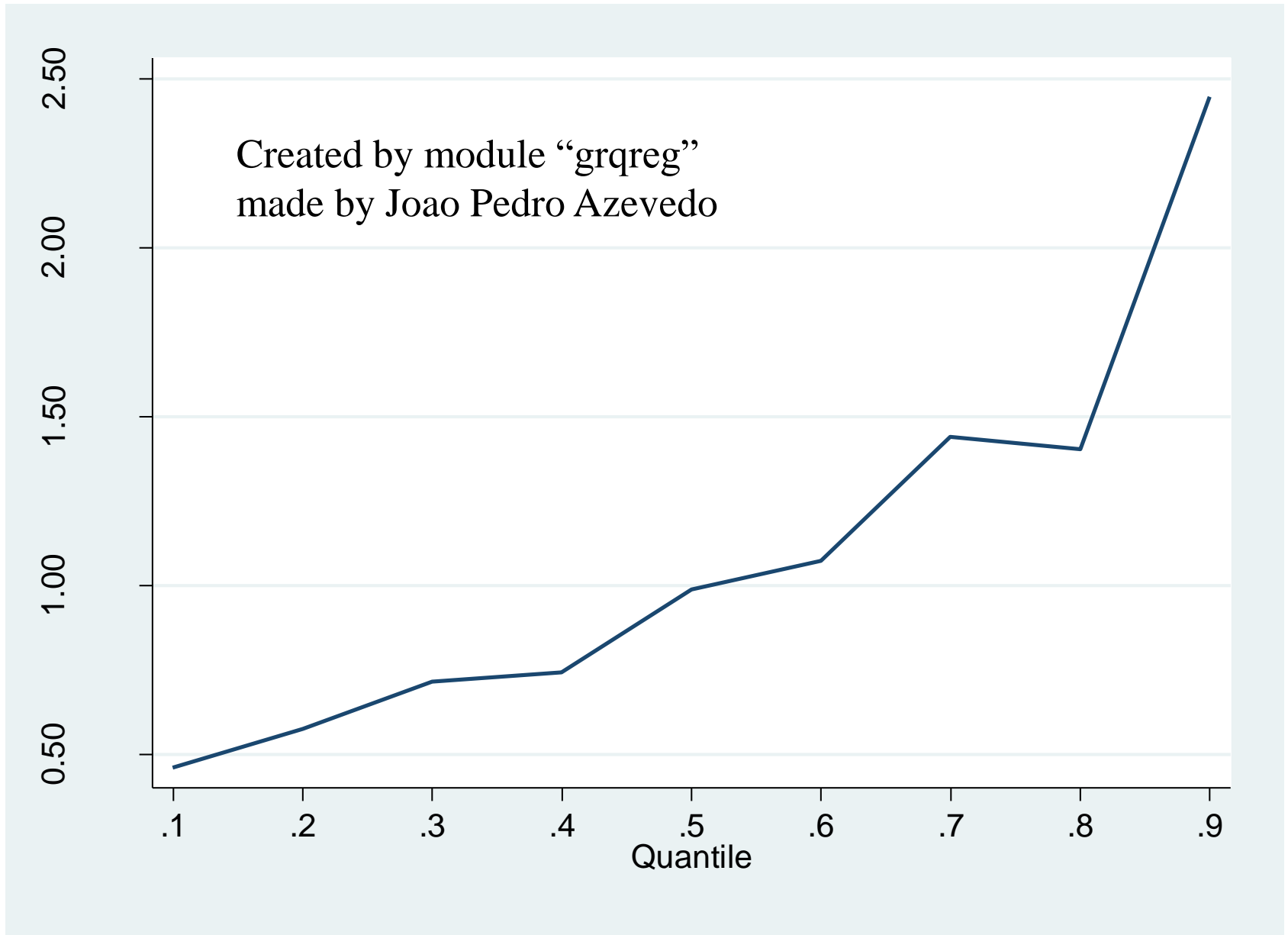
Explanatory variables	Target quantiles								
	.10	.20	.30	.40	.50	.60	.70	.80	.90
Extraversion (centered)	0.46 ***	0.58 ***	0.72 ***	0.74 ***	0.99 ***	1.07 ***	1.44 ***	1.40 ***	2.45 ***
Age (centered)	0.07 **	0.06 +	0.08 **	0.06	0.07	0.13 +	0.17 *	0.21 *	0.24 *
Gender (Female=1, Male=0)	0.71	0.25	-0.09	-0.86	-1.10	-0.81	-2.43	-6.13 *	-8.88 *
Education years (centered)	-0.09	0.00	0.12	-0.14	-0.05	-0.10	-0.38	-0.09	-0.41
Constant	2.85 ***	4.71 ***	6.40 ***	8.30 ***	10.54 ***	13.06 ***	17.45 ***	23.98 ***	32.76 ***
Variance explained	.07	.07	.06	.06	.06	.06	.08	.08	.10

SEs are calculated by a bootstrap method. The number of replications is 500.

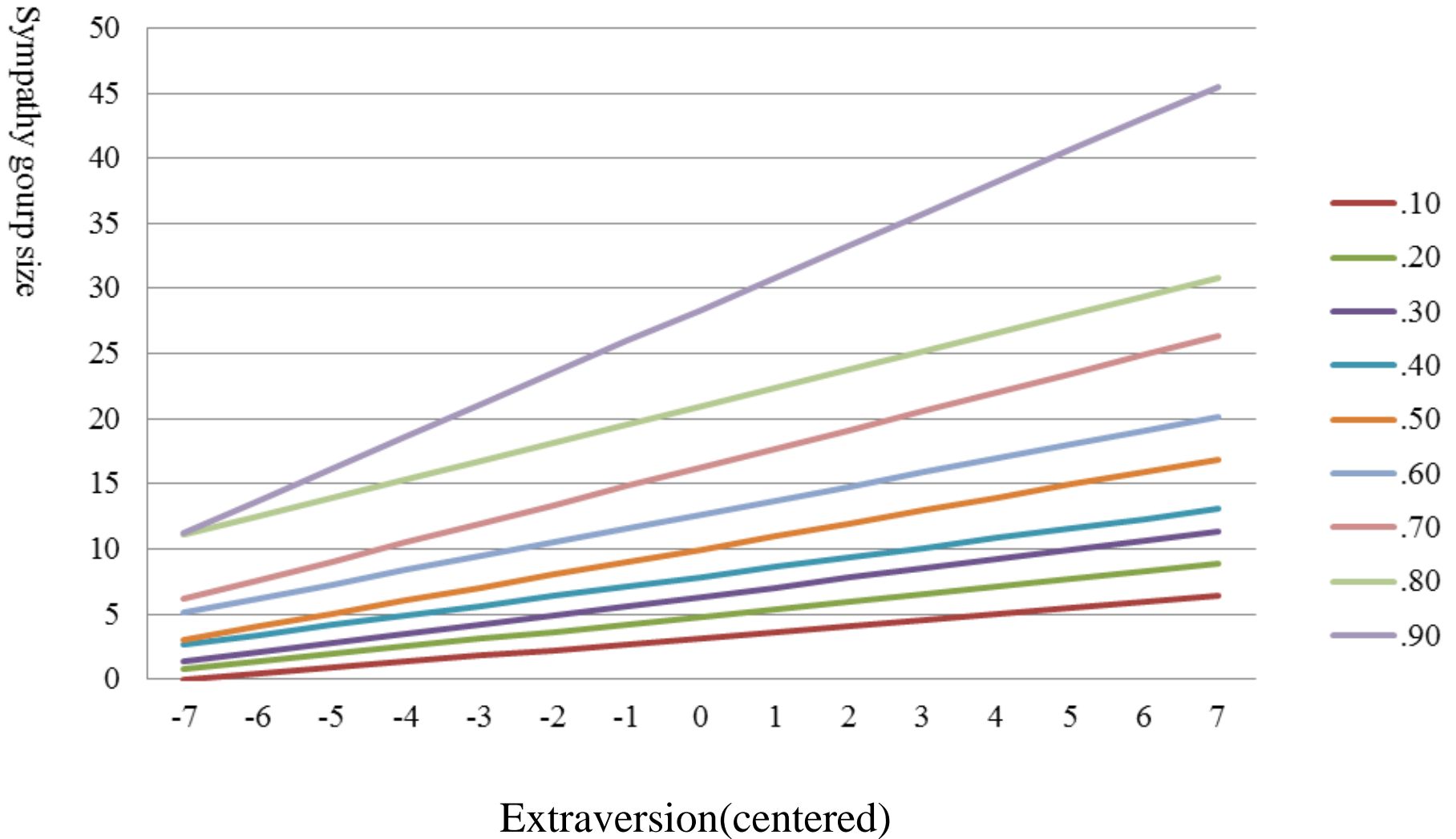
\*\*\*:  $p < .001$ , \*\*:  $p < .01$ , \*:  $p < .05$ , +:  $p < .10$

N=324

# The Comparison of Slopes of Extraversion



# The Relationships of Nine Quantiles and Extraversion



# Results

- The distribution of the size of the sympathy group shifts toward the right as extraversion increases.
- The change, however, is not parallel.
- The right part of the distribution is influenced to a greater extent than the left part.
  - That is, the distribution becomes more right-skewed as extraversion increases.



# Interpretation

- Extraverts cannot always invest their resources, including their free time, in social activities.
- A minimal social relationship is independent of extraversion because it is formed and maintained by social foci, not individual commitment.
  - The upper limit of the network size may be determined by extraversion. However, the actual network size is constrained by many factors.
- We cannot determine these implications using standard regression analysis.

# In the next version I would like...

- Mixed effect/time series quantile regression!
- Graphics of the change of predicted probability density function according to the change of explanatory variables.
- Increased computation time!!
  - bsqreg and sqreg use only 1.5CPUs on my MP4 version.
- Anyway, I hope that more sociological and psychological researchers begin using quantile regression rather than OLS regression.