Topic Chosen _____

Group members

1. _____ id._____
2. _____ id._____
3. _____ id._____
4. _____ id._____
5. _____ id._____

Part 1: Replicate past seminar paper using only the original dataset

- Display the summary statistics table
- Display the result table(s) – The result tables should be in the standard format. You should export your results using the outreg2 command. Then, copy and paste the results in your word document. Alternatively, you can edit the look of your result table in a spreadsheet (excel) first.

Example 1: a standard format of the result table

Table 4. *Fixed-Effects Regressions of Women's Log Wages by Interactions Among Number of Children, Race, and Marital Status*

	White	African American	Hispanic
Variables	*b* (*SE*)	*b* (*SE*)	*b* (*SE*)
1 child	−0.02* (0.01)	0.01 (0.02)	0.00 (0.02)
2 children	−0.08** (0.01)	−0.00 (0.02)	−0.01 (0.02)
3 children	−0.12** (0.02)	−0.04† (0.02)	−0.01 (0.03)
4+ children	−0.09** (0.03)	−0.03† (0.02)	0.05 (0.04)
Never married	−0.08** (0.01)	−0.07** (0.02)	−0.04† (0.02)
Divorced, separated, widowed	−0.02 (0.01)	−0.04† (0.02)	−0.01 (0.03)
Interaction variables			
Never Married × 1 Child	0.04 (0.03)	0.03 (0.02)	0.05 (0.03)
Never Married× 2 Children	0.01 (0.05)	0.01 (0.02)	0.05 (0.04)
Never Married × 3 Children	−0.05 (0.11)	0.06† (0.03)	0.09 (0.06)
Never Married × 4+ Children	−0.08 (0.13)	0.15** (0.05)	0.09 (0.08)
Divorced, Separated, Widowed × 1 Child	0.02 (0.02)	0.02 (0.03)	0.02 (0.03)
Divorced, Separated, Widowed × 2 Children	0.05* (0.02)	0.01 (0.03)	0.01 (0.04)
Divorced, Separated, Widowed × 3 Children	0.11** (0.03)	0.10** (0.03)	0.02 (0.05)
Divorced, Separated, Widowed × 4+ Children	0.16* (0.06)	0.07 (0.04)	−0.09 (0.06)
Constant	0.15 (0.18)	0.35 (0.22)	−0.25 (0.62)
R^2	0.68	0.68	0.67
N of person-years	38,162	17,472	11,220
N of persons	3,503	1,471	955

Note: Models use survey weights and include job-related variables reported in Table 3 as well as control variables for education, age, age-squared, and indicator variables for survey years.

†$p < .10$. *$p < .05$. **$p < .01$.

Example 2: another standard format of the result table (you can choose to follow either example 1 or 2, but please stick with one format throughout your report).

Table 5: Log earnings regressions, Men

	OLS coefficient estimates		Fixed-effect coefficient estimates	
Controls	**1**	**2**	**1**	**2**
Married	.339***	.287***	.139***	.042***
	(.004)	(.011)	(.002)	(.005)
Married*Cohort 1946-55		.067***		.089***
		(.014)		(.006)
Married*Cohort 1956-65		.053***		.110***
		(.013)		(.006)
Married*Cohort 1966-75		.075***		.163***
		(.016)		(.007)

Notes: N=1,617,700 (to nearest 100). Dependent variable is log annual earnings from SSA records for years 1978-2011. Regressions also include controls for year, education, dummies indicating if race is African-American, and indicating if ethnicity is Hispanic, main effects for birth cohort, and a quartic in age.

Part 2: Replicate past seminar paper using the original dataset + additional observations that your group collected.

- Display the summary statistics table
- Display the result table(s)
- Discuss whether the results in part 1 and part 2 are the same. Of course, all the numbers will be different. You do not need to discuss how each pair of number is different. However, you will have to discuss whether results in part 1 and part 2 will give you the same conclusion about the direction and magnitude of impacts of each explanatory variable.

Part 3: Discussion on this research paper in general (1-3 pages)

- Do you think the results are reliable?
  - What are the other variables that should be collected to reduce the potential biasness of the OLS estimators?
- What should the researcher do to improve the quality of their research in general?
  - Hint: your comments can cover a wide range of issues. For example, whether the research question is appropriate, whether the estimation satisfies all the OLS assumptions, whether the author have done a good enough job to obtain unbiased estimator(s) of the slope

coefficient(s), whether there is enough variation in the original dataset collected by the researcher (e.g. whether it covers a wide range of population, income group, age group, etc.).

- You can include some tabulations or histograms of the data if necessary.