

Please answer these questions as completely as possible. Make sure you answer all parts of the question. The exam is open-note, open-book, open-Internet. If you do consult sources, do not directly quote them, but paraphrase and **synthesize** (and provide citations). You do not need to provide citations for my slides or Wooldridge, but do not just copy-and-paste from us either. Do not discuss this exam with anyone until after the due date. Thanks!

**The exam is due on Courseworks under Assignments on Wednesday, Oct. 24th, 1pm.**

*Part 1-- 43 points*

Using the 2016 GSS, a researcher wanted to answer the following question: What influences what people say is the the ideal number of children for a family to have? The outcome is *Ideal Number of Children*, which represents the number of children in a family a respondent would ideally want, ranging from 0 to 7 children.

In Model 1, she predicts this outcome only using the number of actual children a respondent has, ranging from 0 to 8 children (*Number of Children*).

In Model 2, she further includes the number of siblings the respondent has, ranging from 0 to 8 siblings (*Number of Siblings*), and the respondent's age, which ranges from 18 to 89 (*Age*).

In Model 3, she further includes measures for the socioeconomic status (SES) of the respondent: a series of indicator variables for the highest degree the respondent achieved, with "Less than High School" as the reference category, then ascending to *High School Degree*, then *Some College*, then a *BA Degree* and ultimately, *Graduate Degree*; and the natural log of total family income, ranging from 5.46 to 11.79 (*Ln(Income)*).

Lastly, in Model 4, she wants to consider if a person's sense of their own past might affect their ideal number of children. So she includes a measure of the respondent's sense of how much their standard of living has improved now compared with their parents then: the answers range from much worse (1), somewhat worse (2), about the same (3), somewhat better (4), or much better (5) than their parents was (*Improved Standard of Living*).

**Table 1: Summary Statistics for Part I, GSS 2016**

Variable	Obs.	Mean	Std. Dev.	Min	Max
Ideal Number of Children	755	2.57	0.92	0	7
Number of Children	755	1.65	1.58	0	8
Number of Siblings	755	3.39	2.39	0	8
Age	755	47.00	17.87	18	89
Less Than High School (reference)	755	0.10	0.30	0	1
High School Degree	755	0.54	0.50	0	1
Some College	755	0.08	0.27	0	1
BA Degree	755	0.18	0.38	0	1
Graduate Degree	755	0.10	0.30	0	1
Ln(Income)	755	9.91	1.12	5.46	11.79
Improved Standard of Living	755	2.37	1.19	1	5

**Table 2: Models Predicting Ideal Number of Children, GSS 2016**

	Model 1	Model 2	Model 3	Model 4
Number of Children	0.12*** (0.02)	0.13*** (0.02)	0.13*** (0.02)	0.13*** (0.02)
Number of Siblings		0.08*** (0.01)	0.08*** (0.01)	0.07*** (0.01)
Age		-0.01*** (0.00)	-0.01*** (0.00)	-0.01*** (0.00)
High School Degree			-0.07 (0.11)	-0.05 (0.11)
Some College			-0.11 (0.15)	-0.11 (0.15)
BA Degree			-0.12 (0.14)	-0.11 (0.14)
Graduate Degree			-0.04 (0.16)	-0.02 (0.16)
Ln(Income)			-0.07* (0.03)	-0.08** (0.03)
Improved of Standard of Living				0.06* (0.03)
Intercept	2.39*** (0.05)	2.46*** (0.10)	3.2*** (0.32)	3.53*** (0.35)
N	755	755	755	755
R-sq	0.04	0.10	0.11	0.12
Standard errors in parentheses				
* p<0.05 ** p<0.01 *** p<0.001				

1. (a) In Model 1, interpret the coefficient on *Number of Children* (8 points). (b) Is it statistically significant and how do you know? Please talk about t-statistics in your answer. (9 points) (c) In Model 1, does the correlation (or R-squared) between *Ideal Number of Children* and *Number of Children* seem low or high to you? Based on your prior expectation, give one reason for why you think that relationship is either low or high (4 points).
2. In Model 1, interpret the constant/intercept in this model-- i.e., to whom does it refer? Also note statistical significance. (10 points)
3. (a) In Model 2, interpret the coefficient on *Age*. Note statistical significance. (8 points). (b) In Model 2, *Age* has a negative relationship with *Ideal Number of Children*. Does this relationship make sense to you? Give me one explanation for why you think *Age* would have a negative relationship with *Ideal Number of Children* (4 points).



Part 2-- 57 points

Another researcher wanted to answer a different question using the 2016 GSS. Almost 80% of respondents to the GSS are now given a monetary incentive to take the survey; it is a variable called *Fee in \$*, ranging from \$5 to \$75. The question is: What determines the amount of a fee someone will get, if they request money to complete the survey?

In Model 1, the researcher predicts the size of the incentive (*Fee in \$*) as a function of the total family income of the respondent, measured in 1000s of dollars, which ranges from 2.34 to 131.68 (*Income in 1000s*).

In Model 2, the researcher predicts the natural log of the incentive, ranging from 1.61 to 4.32 ( $\ln(\text{Fee})$ ), as a function of the natural log of the total family income ( $\ln(\text{Income})$ ), which ranges from 5.46 to 11.79.

In Model 3, the researcher includes indicator variables for the potential work statuses of the respondent, with the reference group being "Full Time Work," and the other options being: *Working Part-Time*, *Not Working Currently*, *Laid Off/Unemployed*, *Retired*, *In School*, *"Keeping House"*, and *Miscellaneous Work Status*.

In Model 4, the researcher considers one more factor that might affect how much of an incentive the respondent might get: how physically attractive the respondent was, as determined by the interviewer (*How Attractive Respondent Is (Interviewer Assessment)*), ranging from very unattractive (1), unattractive (2), about average (3), attractive (4), very attractive (5).

**Table 3: Summary Statistics for Part II**

Variable	Obs	Mean	Std. Dev.	Min	Max
Fee in \$	1773	55.43	19.32	5.00	75.00
Income (1000s)	1773	30.66	29.98	2.34	131.68
Ln(Fee)	1773	3.93	0.45	1.61	4.32
Ln(Income)	1773	9.84	1.15	5.46	11.79
Working Full-Time (reference)	1773	0.49	0.50	0	1
Working Part-Time	1773	0.12	0.32	0	1
Not Working Currently	1773	0.02	0.15	0	1
Laid Off/Uemployed	1773	0.05	0.22	0	1
Retired	1773	0.16	0.37	0	1
In School	1773	0.03	0.16	0	1
"Keeping House"	1773	0.10	0.30	0	1
Miscellaneous Work Status	1773	0.03	0.16	0	1
How Attractive Respondent Is (Interviewer Assessment)	1773	3.36	0.75	1	5

**Table 4: Models Predicting Amount of Incentive Respondents Got to Take the Survey, GSS 2016**

	<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>	<b>Model 4</b>
	(Fee in \$)	Ln(Fee)	Ln(Fee)	Ln(Fee)
Income (1000s)	0.09*** (0.02)			
Ln(Income)		0.06*** (0.01)	0.05*** (0.01)	0.05*** (0.01)
Working Part-Time			0.03 (0.04)	0.03 (0.04)
Not Working Currently			0.03 (0.07)	0.03 (0.07)
Laid Off/Uemployed			-0.09 (0.05)	-0.09 (0.05)
Retired			-0.02 (0.03)	-0.01 (0.03)
In School			-0.11 (0.07)	-0.12 (0.07)
"Keeping House"			-0.08* (0.04)	-0.08* (0.04)
Miscellaneous Work Status			-0.26*** (0.07)	-0.25*** (0.07)
How Attractive Respondent Is (Interviewer Assessment)				0.03* (0.01)
Intercept	52.6*** (0.65)	3.30*** (0.092)	3.43*** (0.10)	3.36*** (0.11)
N	1773	1773	1773	1773
R-sq	0.02	0.03	0.04	0.04
Standard errors in parentheses				
* p<0.05 ** p<0.01 *** p<0.001				

6. Interpret the coefficient on *Income (1000s)* in Model 1. Note statistical significance. (6 points). Interpret the coefficient on *Ln(Income)* in Model 2. Note statistical significance. (9 points). Why do economists often prefer the specification in Model 2 over the specification in Model 1? Give two reasons. (8 points).

7. Interpret the coefficient on "*Miscellaneous Work Status*" in Model 3. Note statistical significance. (8 points)

8. The researcher theorized that maybe one reason people would expect a larger incentive amount is if they were not home as much (because they are working), they might feel their time was more valuable, even if we control for overall income levels of the family. He included a series of indicators for work status to test this idea. Then he ran a partial F-test on the whole set of *Work Status* dummy variables ("*Part-Time Work*" through "*Misc. Work Status*"). The result was an F-statistic of 3.48 ( $p < 0.001$ ). (a) Explain what a partial F-test tests for (9 points), and (b) explain whether, from a statistical standpoint, this F-test indicates that the researcher made a good decision to include the *Work Status* dummy variables. (8 points)

9. Interpret the coefficient on *How Attractive Respondent Is (Interviewer Assessment)* in Model 4. Note statistical significance. (5 points). What conclusion do you draw from the sign and statistical significance of this variable and its relationship to the size of incentive given to the respondent? Provide one possible conclusion. (4 points)