10

Choosing How to Present Statistical Results

PROBLEM SET

Answer questions 1 through 3 using the information in table 10A .

Table 10A. Estimated coefficients and standard errors from a model of cumulative grade point average by own SAT scores and roommate's SAT scores, Williams College classes of 1999–2001

	Student's own combined math & verbal SAT score		
	Lowest 15%	Middle 70%	Top 15%
Own verbal SAT score/100	0.205	0.199	0.118
	(0.039)	(0.015)	(0.055)
Own math SAT score/100	0.065	0.112	0.045
	(0.036)	(0.017)	(0.051)
Race (ref. = white)			
Black	–0.181	–0.386	–0.800
	(0.046)	(0.053)	(0.059)
Hispanic	–0.036	–0.254	–0.050
	(0.059)	(0.046)	(0.274)
Native American	–0.238	0.212	dropped
	(0.169)	(0.168)	
Not a U.S. citizen	0.076	0.126	0.055
	(0.091)	(0.055)	(0.066)
Asian	0.210	–0.065	–0.201
	(0.120)	(0.026)	(0.047)
Female	0.262	0.103	0.107
	(0.038)	(0.016)	(0.028)
Roommate's verbal SAT score/100	0.006	0.043	–0.013
	(0.025)	(0.012)	(0.021)
Roommate's math SAT score/100	–0.038	–0.021	0.030
	(0.028)	(0.012)	(0.022)
Sample size	450	2,072	629
R^2	0.41	0.27	0.21

Source: Adapted from David A. Zimmerman, "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment," *Review of Economics and Statistics* 85.1 (2003): 9–23, table 4.

1. For the estimated coefficient on female gender among students with combined SATs in the lowest 15%:
   a. What is the *t*-statistic?
   b. What is the 95% confidence interval?
   c. What is the 99% confidence interval?
   d. What is the *p*-value based on a 2-tailed test?
   e. If * denotes $p < 0.05$ and ** denotes $p < 0.01$, what symbol would accompany the "female" coefficient?

2. Among students in the middle 70% of combined SAT scores, which of the following differences in GPA are statistically significant?
   a. That between black and white students
   b. That between black and Hispanic students
   c. That between Hispanic and Native American students
   d. What additional information (if any) do you need to conduct a formal statistical test for these differences?

3. Answer the following questions using the information in table 10A.
   a. Three models are shown in table 10A. How do they differ? How can you tell from the table?
   b. Is the relationship between gender and GPA statistically significantly different across categories of own combined SAT score?
   c. What additional information (if any) do you need to conduct a formal statistical test for this difference?

   Answer questions 4 through 8 using the information in table 10B.1.

4. What are the lower and upper 90% confidence limits for 1998 median income for all households?
5. Is the change in real household income between 1998 and 1999 statistically significant at $p < 0.10$:
   a. For all households?
   b. For family households?
   c. For nonfamily households?

6. What is the standard error associated with the 1998 estimate of median income for nonfamily households with a female householder? Explain how you calculated it.

7. Calculate 95% confidence intervals around estimated median income for each household type in table 10B.1 and show the results in a new table. Hints: Use the critical value for $p < 0.10$ based on a large sample to calculate the standard error of each estimate. Then multiply the standard error by 1.96 to obtain the 95% CI. A spreadsheet vastly simplifies these calculations.

Table 10B.1. Median income (constant 1999$) by type of household, United States, 1998 and 1999

	1998		1999	
Type of Household	Median income	90% confidence interval (+/–)	Median income	90% confidence interval (+/–)
Family households	48,517	419	49,940	449
Married-couple families	55,475	541	56,827	502
Female householder, no husband present	24,932	669	26,164	594
Male householder no wife present	40,284	1,670	41,838	1,311
Nonfamily households	23,959	477	24,566	444
Female householder	19,026	472	19,917	454
Male householder	31,086	572	30,753	568
All households	39,744	387	40,816	314

Source: U.S. Census Bureau, *Current Population Report*s, P60-209, *Money Income in the United States: 1999* (Washington, DC: U.S. Government Printing Office), table A.

8. Create a table that shows change in median income for each household type between 1998 and 1999, denoting differences that are statistically significant at $p < 0.10$ with a dagger.

Answer questions 9 and 10 using the information in table 10C.

9. For the estimated coefficient on "ever-married," calculate:
   a. The test statistic (name it)
   b. The *p*-value
   c. The 95% confidence interval for the coefficient (e.g., the 95% CI around the log-odds point estimate)

10. Revise table 10C to report odds ratios with associated 95% confidence intervals and symbols to denote statistical significance instead of log-odds and standard errors.

Table 10C. Estimated log-odds of first trip to the United States, Men, 1987–1998 Mexican Migration Project

	Log-odds	Standard error
Demographic background		
Age (years)	–0.003	0.02
Age-squared	–0.001	0.0002
Ever married	–0.09	0.06
Number of minor children in household	0.01	0.01
Human capital		
Years of education	–0.04	0.006
Months of labor-force experience	–0.002	0.0007
Social capital in the family		
Parent a prior U.S. migrant	0.51	0.05
Siblings prior U.S. migrants	0.36	0.02
Social capital in the community		
Migration prevalence ratio[a]		
0–4	–0.99	0.15
5–9	–0.09	0.12
(10–14)		
15–19	0.35	0.10
20–29	0.57	0.13
30–39	0.95	0.15
40–59	0.74	0.19
60 or more	0.34	0.15
Intercept	–3.31	0.26
– 2 log likelihood	23,369.2	
Df	26	

Source: Adapted from Elizabeth Fussell and Douglas S. Massey, "The Limits to Cumulative Causation: International Migration from Mexican Urban Areas," *Demography* 41.1 (2004): 151–71. Table 2, http://muse.jhu.edu/journals/demography/v041/41.1fussell.pdf.

Note: Model also includes controls for occupational sector, internal migratory experience, community characteristics, and Mexican economic and U.S. policy context.

[a] The migration prevalence ratio = the number of people aged 15+ years who had ever been to the U.S./the number of people aged 15+ years ×100.